B44

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/12, C07K 14/47, C12Q 1/68, C12N 15/63, A61K 38/17, C07K 16/18

A1

(11) International Publication Number: WO 00/26372

(43) International Publication Date: 11 May 2000 (11.05.00)

(21) International Application Number: PCT/US99/26177

(22) International Filing Date: 5 November 1999 (05.11.99)

(30) Priority Data:
60/172,248 5 November 1998 (05.11.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/172,248 (CIP)
Filed on 5 November 1998 (05.11.98)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue, #30, Mountain View, CA 94041 (US). GORGONE, Gina, A. [US/US]; 1253 Pinecrest Drive, San Francisco, CA 94132 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MYOSIN HEAVY CHAIN HOMOLOG

```
            10          19          28          37          46          55
5'TGAT GCT CTG GGC TGT CTT CAC ACT TCA TTT GGG TTT CCT GCT TGC TCT GAG CTC

            64          73          82          91         100         109
TAC AGG GGA ATG GGG TAG AGA TGG GAG CCA CCT TGG GTG GAG GGT GGG GAA GGT

           118         127         136         145         154         163
ATG TTC TGC CCA CCA CAG GTG TCA TGC TCA CTC AGC CTG ATG CCC AGG CTG CCA
M   F   C   P   P   Q   V   S   C   S   L   S   L   M   P   R   L   P

           172         181         190         199         208         217
AGT ATA AGG CAT TGG CAG GGG CCC AGC CAC CCT GGG TTC CTT GGT CCC CTA TTC
S   I   R   H   W   Q   G   P   S   H   P   G   F   L   G   P   L   F

           226         235         244         253         262         271
CCC ATC TGC TCC CTG CAG TGG CCC CAT GGG TTC TCT GCC ATC TTC CCA GGC CTG
P   I   C   S   L   Q   W   P   H   G   F   S   A   I   F   P   G   L

           280         289         298         307         316         325
CTG GAT GTG TAT GGA TTT GAA TCA TTT CCT GAC AAC AGT CTG GAA CAG TTG TGC
L   D   V   Y   G   F   E   S   F   P   D   N   S   L   E   Q   L   C

           334         343         352         361         370         379
ATC AAC TAC GCC AAT GAG AAG CTG CAG CAG CAT TTT GTG GCT CAC TAC CTA AGG
I   N   Y   A   N   E   K   L   Q   Q   H   F   V   A   H   Y   L   R
```

(57) Abstract

The invention provides a human myosin heavy chain homolog (MHCH) and polynucleotides which identify and encode MHCH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of MHCH.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

MYOSIN HEAVY CHAIN HOMOLOG

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of a myosin heavy chain homolog and to the use of these sequences in the diagnosis, treatment, and prevention of heart and skeletal muscle disorders; developmental disorders; and cell proliferative disorders including cancer.

BACKGROUND OF THE INVENTION

Myosins are actin-activated ATPases, found in eukaryotic cells, that couple hydrolysis of ATP with motion. Myosin provides the motor function for muscle contraction and intracellular movements such as phagocytosis and rearrangement of cell contents during mitotic cell division (cytokinesis). Myosins are composed of one or two heavy chains and associated light chains. Myosin heavy chains contain an amino-terminal motor or head domain, a neck that is the site of light-chain binding, and a carboxy-terminal tail domain. Conventional myosins, such as those found in muscle tissue, are composed of two myosin heavy-chain subunits, each associated with two light-chain subunits that bind at the neck region and play a regulatory role. Unconventional myosins, believed to function in intracellular motion, may contain either one or two heavy chains and associated light chains. There is evidence for about 25 myosin heavy chain genes in vertebrates, more than half of them unconventional. Recently the myosins have been divided into 11 classes.

The heavy myosin chain head domain ends in an amino acid sequence which is conserved in most myosins. The neck domains of most myosin heavy chains (MyHC) consist of a variable number of motifs with a conserved sequence believed to be the site for light-chain binding. Calmodulin or calmodulin-like proteins function as light chains. An unexpected degree of variation has been observed in the tail domains of different myosins. Several unconventional myosins contain domains associated with signal transduction (Mooseker, M. et al. (1995) Annu. Rev. Cell Dev. Biol. 11:633-675).

Disorders of myosin function are involved in a variety of human diseases including muscle disorders, developmental disorders, and cancer. Two forms of myosin heavy chain (alpha and beta) have been observed in the mammalian ventricular myocardium. The speed with which the heart contracts is related to their relative expression, which suggests that increased alpha MyHC expression may be therapeutic in cardiovascular disease. Decline in atherosclerosis resistance with age has been related to downregulation of non-muscle MyHC (Amore, B. et al. (1996) J. Vasc. Res. 33:442-453). Aging has also been related to decreased class II MyHC expression and an increase in Class I MyHC expression (Larsson, L. et al. (1997) Acta Physiol. Scand. 159:81-89). Mutations in genes coding for

the beta-myosin heavy chain have been related to hypertrophic cardiomyopathy (Marian, A.J. et al. (1998) J. Cardiovasc. Electrophysiol. 9:88-99).

Growth and muscle defects are observed as a result of defects in MyHC expression (Acakpo-Satchivi, L.J. (1997) J. Cell Biol. 139:1219-1229). Abnormal MyHC expression has been observed in muscular dystrophy (Tidyman, W.E. et al. (1997) Dev. Dyn. 208:491-504). Deafness associated with disruption in organization of the hair cells of the inner ear results from a mutant unconventional-myosin gene (Probst, F.J. et al. (1998) Science 280:1444-1447).

Modulation of MyHC has been implicated in breast cancer (Ohyabu, I. et al. (1998) Pathol. Int. 48:433-439). The chromosomal aberration associated with acute myeloid leukemia produces a protein that contains the rod domain of the smooth muscle MyHC molecule (Tanaka, Y. et al. (1998) Oncogene 17:699-708). Myosin V is associated with transport of epidermal melanocytes. The S91 mouse melanoma cell line shows a defect in transport of melanosomes (Brown, D.A. et al. (1998) J. Invest. Dermatol.110:428-437).

The discovery of a new myosin heavy chain homolog and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of heart and skeletal muscle disorders; developmental disorders; and cell proliferative disorders including cancer.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human myosin heavy chain homolog, referred to as "MHCH." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also includes a polypeptide comprising an amino acid sequence that differs by one or more conservative amino acid substitutions from an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes

under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing a polynucleotide of the invention under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of MHCH, the method comprising administering to a subject in need

of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of MHCH, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

Figures 1A, 1B, 1C, 1D, 1E, and 1F show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of MHCH. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, 2J, and 2K show the amino acid sequence alignment among MHCH (Incyte Clone ID 1929760; SEQ ID NO:1), Caenorhabditis elegans myosin I heavy chain (GI 1279777; SEQ ID NO:3), and Helianthus annuus (sunflower) unconventional myosin I heavy chain (GI 2444174; SEQ ID NO:4), produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

Table 1 shows the tools, programs, and algorithms used to analyze MHCH, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be

used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"MHCH" refers to the amino acid sequences of substantially purified MHCH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of MHCH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MHCH either by directly interacting with MHCH or by acting on components of the biological pathway in which MHCH participates.

An "allelic variant" is an alternative form of the gene encoding MHCH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding MHCH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as MHCH or a polypeptide with at least one functional characteristic of MHCH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding MHCH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding MHCH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MHCH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of MHCH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine.

Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of MHCH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MHCH either by directly interacting with MHCH or by acting on components of the biological pathway in which MHCH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind MHCH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the

complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic MHCH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding MHCH or fragments of MHCH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded

as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of MHCH or the polynucleotide encoding MHCH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues

in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:2 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:2, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:2 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:2 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:2 and the region of SEQ ID NO:2 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1 is encoded by a fragment of SEQ ID NO:2. A fragment of SEQ ID NO:1 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1. For example, a fragment of SEQ ID NO:1 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1. The precise length of a fragment of SEQ ID NO:1 and the region of SEQ ID NO:1 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer

to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 11
Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence,

for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour.

Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of MHCH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of MHCH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding MHCH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding MHCH, or fragments thereof, or MHCH itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or

synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the

reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of a new human myosin heavy chain homolog (MHCH), the polynucleotides encoding MHCH, and the use of these compositions for the diagnosis, treatment, or prevention of heart and skeletal muscle disorders; developmental disorders; and cell proliferative disorders including cancer.

Nucleic acids encoding the MHCH of the present invention were identified in Incyte Clone 1929760H1 from the colon tumor cDNA library (COLNTUT03) using a computer search for nucleotide and/or amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1929760H1 and 1929760F6 (COLNTUT03), 2418744F3 (HNT3AZT01), 3229696X11F1 (COTRNOT01), 3344480F6 (SPLNNOT09), 401389H1 (TMLR3DT01), 5111681H1 (ENDITXT01), 1451483H1 (PENITUT01), and shotgun sequence SBCA04642F1.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C, 1D, 1E, and 1F. MHCH is 612 amino acids in length and has 7 potential casein kinase II phosphorylation sites at residues S62, T146, T221, S280, S323, S390, and T546; and 6 potential protein kinase C phosphorylation sites at residues S19, S140, S303, T441, S555, and S563. The sequence from T383 through M387 of MHCH is 80% identical to the conserved sequence found at the end of myosin head domains. MHCH contains two possible light-chain binding sites. The first, from I410 through E421, contains 4 out of 6 conserved residues and the second, from I432 through K443, contains 5 out of 6 conserved residues. PFAM analysis shows that MHCH shares homology with a myosin head domain from residue F51 to residue

L314. PRINTS analysis shows that MHCH shares homology with a myosin heavy chain signature motif from residue F51 to K79 and from residue F105 to C133. MHCH has a possible transmembrane motif from residue W506 to P535. As shown in Figures 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, 2J, and 2K, MHCH has chemical and structural similarity with Caenorhabditis elegans myosin I heavy chain (GI 1279777; SEQ ID NO:3), and Helianthus annuus unconventional myosin heavy chain (GI 2444174; SEQ ID NO:4). MHCH and myosin I heavy chain share 23.2% identity, and in particular they share 39% identity from residue F51 to residue L314 of MHCH. MHCH and unconventional myosin I share 22.4% identity, and in particular they share 38% identity from residue F51 to residue L314 of MHCH. A fragment of SEQ ID NO:2 from about nucleotide 122 to about nucleotide 166 is useful in hybridization or amplification technologies to identify SEQ ID NO:2 and to distinguish between SEQ ID NO:2 and a related sequence. The encoded polypeptide is useful, for example, as an antigenic polypeptide. Northern analysis shows the expression of this sequence in various libraries, at least 65% of which are associated with cell proliferation or cancer, at least 34% of which are associated with the immune response, at least 24% of which are associated with gastrointestinal tissue, at least 24% of which are associated with reproductive tissue, at least 13% of which are associated with hematopoietic/immune tissue, at least 10% are associated with musculoskeletal tissue, and at least 10% are associated with nervous tissue.

The invention also encompasses MHCH variants. A preferred MHCH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the MHCH amino acid sequence, and which contains at least one functional or structural characteristic of MHCH.

The invention also encompasses polynucleotides which encode MHCH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes MHCH.

The invention also encompasses a variant of a polynucleotide sequence encoding MHCH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding MHCH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising the sequence of SEQ ID NO:2 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence consisting of SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MHCH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the

genetic code. a multitude of polynucleotide sequences encoding MHCH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring MHCH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode MHCH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring MHCH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding MHCH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding MHCH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode MHCH and MHCH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding MHCH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is

automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding MHCH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode MHCH may be cloned in recombinant DNA molecules that direct expression of MHCH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express MHCH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter MHCH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding MHCH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, MHCH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of MHCH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New

York NY.)

In order to express a biologically active MHCH, the nucleotide sequences encoding MHCH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding MHCH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding MHCH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding MHCH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MHCH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MHCH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding MHCH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding MHCH can be achieved using a

multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding MHCH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of MHCH are needed, e.g. for the production of antibodies, vectors which direct high level expression of MHCH may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of MHCH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of MHCH. Transcription of sequences encoding MHCH may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding MHCH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses MHCH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of

DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of MHCH in cell lines is preferred. For example, sequences encoding MHCH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding MHCH is inserted within a marker gene sequence, transformed cells containing sequences encoding MHCH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding MHCH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates

expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding MHCH and that express MHCH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of MHCH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MHCH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MHCH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding MHCH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding MHCH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MHCH may be designed to contain signal sequences which direct secretion of MHCH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding MHCH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric MHCH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of MHCH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the MHCH encoding sequence and the heterologous protein sequence, so that MHCH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled MHCH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of MHCH may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein

synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of MHCH may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of MHCH and myosin heavy chain. In addition, the expression of MHCH is closely associated with cell proliferation or cancer, the immune response, gastrointestinal tissue, reproductive tissue, musculoskeletal tissue, and nervous tissue. Therefore, MHCH appears to play a role in heart and skeletal muscle disorders; developmental disorders; and cell proliferative disorders including cancer. In the treatment of disorders associated with increased MHCH expression or activity, it is desirable to decrease the expression or activity of MHCH. In the treatment of disorders associated with decreased MHCH expression or activity, it is desirable to increase the expression or activity of MHCH.

Therefore, in one embodiment, MHCH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MHCH. Examples of such disorders include, but are not limited to, a heart or skeletal muscle disorder such as angina, anaphylactic shock, arrhythmias, asthma, Becker's muscular dystrophy, cardiovascular shock, central core disease, Cushing's syndrome, Duchenne's muscular dystrophy, encephalopathy, epilepsy, hypertension, hypoglycemia, Kearns-Sayre syndrome, lactic acidosis, migraine, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, myocardial infarction, myotonic dystrophy, myocarditis, myoclonic disorder, ophthalmoplegia, pheochromocytoma, and myopathies including cardiomyopathy, centronuclear myopathy, ethanol myopathy, lipid myopathy, mitochondrial myopathy nemaline myopathy, and thyrotoxic myopathy; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, or uterus.

In another embodiment, a vector capable of expressing MHCH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MHCH including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified MHCH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MHCH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of MHCH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MHCH including, but not limited to, those listed above.

In a further embodiment, an antagonist of MHCH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MHCH. Examples of such disorders include, but are not limited to, those heart and skeletal muscle disorders; developmental disorders; and cell proliferative disorders including cancer described above. In one aspect, an antibody which specifically binds MHCH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express MHCH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding MHCH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MHCH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of MHCH may be produced using methods which are generally known in the art. In particular, purified MHCH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind MHCH. Antibodies to MHCH may also be generated using methods that are well known in the art. Such antibodies may include, but are

not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with MHCH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to MHCH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of MHCH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to MHCH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce MHCH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as

disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for MHCH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between MHCH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering MHCH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for MHCH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of MHCH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple MHCH epitopes, represents the average affinity, or avidity, of the antibodies for MHCH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular MHCH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the MHCH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of MHCH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of MHCH-

antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding MHCH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding MHCH may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding MHCH. Thus, complementary molecules or fragments may be used to modulate MHCH activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding MHCH.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding MHCH. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding MHCH can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding MHCH. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding MHCH. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block

translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding MHCH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding MHCH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of MHCH, antibodies to MHCH, and mimetics, agonists, antagonists, or inhibitors of MHCH. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing. compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar

solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of MHCH, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example MHCH or fragments thereof, antibodies of MHCH, and agonists, antagonists or inhibitors of MHCH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind MHCH may be used for the diagnosis of disorders characterized by expression of MHCH, or in assays to monitor patients being treated with MHCH or agonists, antagonists, or inhibitors of MHCH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for MHCH include methods which utilize the antibody and a label to detect MHCH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring MHCH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of MHCH expression. Normal or standard values for MHCH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to MHCH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of MHCH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding MHCH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of MHCH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of MHCH, and to monitor regulation of MHCH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding MHCH or closely related molecules may be used to identify nucleic acid sequences which encode MHCH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding MHCH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the MHCH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the MHCH gene.

Means for producing specific hybridization probes for DNAs encoding MHCH include the cloning of polynucleotide sequences encoding MHCH or MHCH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding MHCH may be used for the diagnosis of disorders associated with expression of MHCH. Examples of such disorders include, but are not limited to, a heart or skeletal muscle disorder such as angina, anaphylactic shock, arrhythmias, asthma, Becker's muscular dystrophy, cardiovascular shock, central core disease, Cushing's syndrome, Duchenne's muscular dystrophy, encephalopathy, epilepsy, hypertension, hypoglycemia, Kearns-Sayre syndrome, lactic acidosis, migraine, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, myocardial infarction, myotonic dystrophy, myocarditis, myoclonic disorder, ophthalmoplegia, pheochromocytoma, and myopathies including cardiomyopathy, centronuclear myopathy, ethanol myopathy, lipid myopathy, mitochondrial myopathy nemaline myopathy, and thyrotoxic myopathy; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, or uterus. The polynucleotide sequences encoding MHCH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered MHCH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding MHCH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding MHCH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding MHCH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of MHCH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding MHCH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding MHCH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding MHCH, or a fragment of a polynucleotide complementary to the polynucleotide encoding

MHCH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of MHCH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding MHCH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding MHCH on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder.

The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, MHCH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between MHCH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with MHCH, or fragments thereof, and washed. Bound MHCH is then detected by methods well known in the art. Purified MHCH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding MHCH specifically compete with a test compound for binding MHCH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with MHCH.

In additional embodiments, the nucleotide sequences which encode MHCH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0621 P, filed November 5, 1998], are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

The COLNTUT03 library was constructed from tumorous sigmoidal colon tissue, obtained from a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated an invasive grade 2 (of 4) adenocarcinoma. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Previous surgeries included cholecystectomy and repair of indirect inguinal hernia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer, and prostate cancer.

The frozen tissue was homogenized and lysed in guanidinium isothiocyanate solution using a Polytron-PT 3000 homogenizer (Brinkmann Instruments, Westbury NY). RNA was isolated as per Stratagene's RNA isolation protocol (Stratagene, La Jolla CA). RNA was extracted twice with acid phenol, precipitated with sodium acetate and ethanol, resuspended in RNase-free water, and treated with DNase. Poly (A+) RNA was isolated using the OLIGOTEX kit (QIAGEN Inc, Valencia CA).

Poly (A+) RNA was used to construct the COLNTUT03 cDNA library according to the recommended protocols in the SUPERSCRIPT plasmid system (Life Technologies). The cDNAs were fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech), and those cDNAs exceeding 400 bp were ligated into the plasmid pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into DH5α competent cells (Life Technologies).

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or

without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 1 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 1 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned

sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:2. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are

usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding MHCH occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in the description of the invention.

V. Extension of MHCH Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:2 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE

and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:2 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ-^{32}P] adenosine

triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the MHCH-encoding sequences, or any parts thereof, are used

to detect, decrease, or inhibit expression of naturally occurring MHCH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of MHCH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the MHCH-encoding transcript.

IX. Expression of MHCH

Expression and purification of MHCH is achieved using bacterial or virus-based expression systems. For expression of MHCH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express MHCH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of MHCH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding MHCH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, MHCH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from MHCH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate

resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified MHCH obtained by these methods can be used directly in the following activity assay.

X. Demonstration of MHCH Activity

The assay for MHCH activity is based upon the ability of MHCH to interact with actinomyosin filaments in vitro (Ho, G. and R.L. Chisholm (1997) J. Biol. Chem. 272:4522-4527). Actin-activated ATPase is assayed in buffer A (10 mM Tris-HCl, pH 7.6, 25 mM KCl, 5 mM $MgCl_2$, 0.1 mM $CaCl_2$, 1 mM ATP), 0-10 μM MHCH, 0-10 μM actin, and 50 μg/ml myosin. Ca^{2+}-activated ATPase is assayed in buffer B (20 mM Tris-HCl, pH 8.0, 500 mM KCl, 10 mM $CaCl_2$, 1 mM ATP), 0-10 μM MHCH, and 50 μg/ml myosin. Reactions are incubated at room temperature for 5 min and then quenched with acid, and the liberated inorganic phosphate (P_i) is quantified following organic extraction.

In vitro motility assays are performed as follows. Myosin is diluted to 200 μg/ml in buffer C (25 mM imidazole, pH 7.4, 25 mM KCl, 4 mM $MgCl_2$, 1 mM EGTA, 10 mM dithiothreitol), applied to a flow cell coated with nitrocellulose, and blocked with buffer C containing 0.5 mg/ml BSA (C/BSA). A solution of phalloidin-labeled actin is perfused followed by 1 mM ATP in C/BSA to remove myosin heads that bind actin in a rigorous fashion. After washing with C/BSA to remove the excess nonfluorescent actin, a solution of rhodamine-phalloidin-labeled actin and MHCH in C/BSA is introduced. Active movement is initiated at room temperature by introducing C/BSA containing 1 mM ATP and oxygen scavenger enzymes. Microscopic images of moving myotubes are tracked for up to 30s, and translocation velocities calculated using the myotube centroids to establish initial and final positions for 2s or 4s samples during the continuous movement.

XI. Functional Assays

MHCH function is assessed by expressing the sequences encoding MHCH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-

based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MHCH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MHCH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MHCH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of MHCH Specific Antibodies

MHCH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the MHCH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-MHCH activity by, for example, binding the peptide or MHCH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purificati n of Naturally Occurring MHCH Using Specific Antibodies

Naturally occurring or recombinant MHCH is substantially purified by immunoaffinity chromatography using antibodies specific for MHCH. An immunoaffinity column is constructed by covalently coupling anti-MHCH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MHCH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of MHCH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MHCH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MHCH is collected.

XIV. Identification of Molecules Which Interact with MHCH

MHCH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MHCH, washed, and any wells with labeled MHCH complex are assayed. Data obtained using different concentrations of MHCH are used to calculate values for the number, affinity, and association of MHCH with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 1 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:2 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression or activity of MHCH, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression or activity of MHCH, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

```
                10        19        28        37        46        55
5'TGAT GCT CTG GGC TGT CTT CAC ACT TCA TTT GGG TTT CCT GCT TGC TCT GAG CTC

                64        73        82        91       100       109
   TAC AGG GGA ATG GGG TAG AGA TGG GAG CCA CCT TGG GTG GAG GGT GGG GAA GGT

               118       127       136       145       154       163
   ATG TTC TGC CCA CCA CAG GTG TCA TGC TCA CTC AGC CTG ATG CCC AGG CTG CCA
   M   F   C   P   P   Q   V   S   C   S   L   S   L   M   P   R   L   P

               172       181       190       199       208       217
   AGT ATA AGG CAT TGG CAG GGG CCC AGC CAC CCT GGG TTC CTT GGT CCC CTA TTC
   S   I   R   H   W   Q   G   P   S   H   P   G   F   L   G   P   L   F

               226       235       244       253       262       271
   CCC ATC TGC TCC CTG CAG TGG CCC CAT GGG TTC TCT GCC ATC TTC CCA GGC CTG
   P   I   C   S   L   Q   W   P   H   G   F   S   A   I   F   P   G   L

               280       289       298       307       316       325
   CTG GAT GTG TAT GGA TTT GAA TCA TTT CCT GAC AAC AGT CTG GAA CAG TTG TGC
   L   D   V   Y   G   F   E   S   F   P   D   N   S   L   E   Q   L   C

               334       343       352       361       370       379
   ATC AAC TAC GCC AAT GAG AAG CTG CAG CAG CAT TTT GTG GCT CAC TAC CTA AGG
   I   N   Y   A   N   E   K   L   Q   Q   H   F   V   A   H   Y   L   R
```

FIGURE 1A

```
          388                 397                 406                 415                 424                 433
GCC CAG CAG GAG GAA TAC GCA GTT GAG GGC CTG GAG TGG TCA TTC ATC AAC TAC
A   Q   Q   E   E   Y   A   V   E   G   L   E   W   S   F   I   N   Y

          442                 451                 460                 469                 478                 487
CAG GAC AAC CAG CCC TGT TTG GAT CTC ATT GAG GGA AGC CCC ATC AGC ATC TGC
Q   D   N   Q   P   C   L   D   L   I   E   G   S   P   I   S   I   C

          496                 505                 514                 523                 532                 541
TCC CTC ATA AAT GAG GAA TGC CGC CTC AAT CGA CCC AGC AGC GCA CGC CAG CTC
S   L   I   N   E   E   C   R   L   N   R   P   S   S   A   R   Q   L

          550                 559                 568                 577                 586                 595
CAG ACA CGC ATT GAG ACT GCC CTG GCA GGC AGC CCC TGC CTG GGC CAC AAT AAG
Q   T   R   I   E   T   A   L   A   G   S   P   C   L   G   H   N   K

          604                 613                 622                 631                 640                 649
CTC AGC CGG GAG CCC AGC TTC ATT GTG GTG CAT TAT GCG GGG CCT GTG CGG TAC
L   S   R   E   P   S   F   I   V   V   H   Y   A   G   P   V   R   Y

          658                 667                 676                 685                 694                 703
CAC ACA GCA GGC CTG GTG GAG AAG AAC AAG GAC CCT ATC CCA CCT GAG CTG ACC
H   T   A   G   L   V   E   K   N   K   D   P   I   P   P   E   L   T

          712                 721                 730                 739                 748                 757
AGG CTC CTG CAG CAA TCC CAG GAC CCC CTG CTC ATG GGG CTG TTT CCT ACT AAC
R   L   L   Q   Q   S   Q   D   P   L   L   M   G   L   F   P   T   N
```

FIGURE 1B

```
          766        775        784        793        802        811
CCC AAA GAG AAG ACC CAG GAG GAA CCC CCT GGC CAG AGC AGG GCC CCT GTG TTG
P   K   E   K   T   Q   E   E   P   P   G   Q   S   R   A   P   V   L

          820        829        838        847        856        865
ACC GTG GTG TCC AAG TTC AAG GCC TCA CTG GAG CAG CTT CTG CAG GTC CTA CAC
T   V   V   S   K   F   K   A   S   L   E   Q   L   L   Q   V   L   H

          874        883        892        901        910        919
AGC ACC ACG CCC CAC TAC ATT CGC TGC ATC AAG CCC AAC AGC CAG GGC CAG GCG
S   T   T   P   H   Y   I   R   C   I   K   P   N   S   Q   G   Q   A

          928        937        946        955        964        973
CAG ACC TTT CTC CAA GAG GAG GTC CTG AGC CAG CTG GAG GCC TGT GGC CTC GTG
Q   T   F   L   Q   E   E   V   L   S   Q   L   E   A   C   G   L   V

          982        991       1000       1009       1018       1027
GAG ACC ATC CAT ATC AGT GCT GCT GGC TTC CCC ATC CGG GTC TCT CAC CGA AAC
E   T   I   H   I   S   A   A   G   F   P   I   R   V   S   H   R   N

         1036       1045       1054       1063       1072       1081
TTT GTA GAA CGA TAC AAG TTA CTA AGA AGG CTT CAT CCT TGC ACA TCC TCT GGC
F   V   E   R   Y   K   L   L   R   R   L   H   P   C   T   S   S   G

         1090       1099       1108       1117       1126       1135
CCC GAC AGC CCA TAT CCT GCC AAA GGG CTC CCT GAA TGG TGT CCA CAC AGC GAG
P   D   S   P   Y   P   A   K   G   L   P   E   W   C   P   H   S   E
```

FIGURE 1C

```
      1144        1153        1162        1171        1180        1189
GAA GCC ACG CTT GAA CCT CTC ATC CAG GAC ATT CTC CAC ACT CTG CCG GTC CTA
E   A   T   L   E   P   L   I   Q   D   I   L   H   T   L   P   V   L

      1198        1207        1216        1225        1234        1243
ACT CAG GCA GCA GCC ATA ACT GGT GAC TCG GCT GAG GCC ATG CCA GCC CCC ATG
T   Q   A   A   A   I   T   G   D   S   A   E   A   M   P   A   P   M

      1252        1261        1270        1279        1288        1297
CAC TGT GGC AGG ACC AAG GTG TTC ATG ACT GAC TCT ATG CTG GAG CTT CTG GAA
H   C   G   R   T   K   V   F   M   T   D   S   M   L   E   L   L   E

      1306        1315        1324        1333        1342        1351
TGT GGG CGT GCC CGG GTG CTG GAG CAG TGT GCC CGC TGC ATC CAG GGT GGC TGG
C   G   R   A   R   V   L   E   Q   C   A   R   C   I   Q   G   G   W

      1360        1369        1378        1387        1396        1405
AGG CGA CAC CGG CAC CGA GAG CAG GAG CGG CAG TGG CGG GCC GTC ATG CTC ATC
R   R   H   R   H   R   E   Q   E   R   Q   W   R   A   V   M   L   I

      1414        1423        1432        1441        1450        1459
CAG GCA GCC ATT CGT TCC TGG TTA ACT CGG AAA CAC ATC CAG AGG CTG CAT GCA
Q   A   A   I   R   S   W   L   T   R   K   H   I   Q   R   L   H   A
```

FIGURE 1D

```
      1468        1477        1486        1495        1504        1513
GCT GCC ACA GTC ATC AAG CGT GCA TGG CAG AAG TGG AGA ATC AGA ATG GCC TGC
A   A   T   V   I   K   R   A   W   Q   K   W   R   I   R   M   A   C

      1522        1531        1540        1549        1558        1567
CTT GCT GCT AAA GAG CTG GAT GGT GTG GAA GAA AAA CAC TTC TCT CAA GCT CCC
L   A   A   K   E   L   D   G   V   E   E   K   H   F   S   Q   A   P

      1576        1585        1594        1603        1612        1621
TGT TCC CTG AGC ACC TCG CCG CTG CAG ACC AGG CTC CTG GAG GCA ATA ATC CGC
C   S   L   S   T   S   P   L   Q   T   R   L   L   E   A   I   I   R

      1630        1639        1648        1657        1666        1675
CTC TGG CCC CTG GGA CTG GTC CTG GCC AAT ACG GCT ATG GGT GTA GGC AGC TTT
L   W   P   L   G   L   V   L   A   N   T   A   M   G   V   G   S   F

      1684        1693        1702        1711        1720        1729
CAG AGG AAA TTA GTG GTC TGG GCT TGC CTC CAG CTC CCC AGG GGC AGC CCC AGT
Q   R   K   L   V   V   W   A   C   L   Q   L   P   R   G   S   P   S

      1738        1747        1756        1765        1774        1783
AGC TAC ACT GTC CAG ACA GCA CAA GAC CAG GCT GGT GTC ACG TCC ATC CGA GCG
S   Y   T   V   Q   T   A   Q   D   Q   A   G   V   T   S   I   R   A

      1792        1801        1810        1819        1828        1837
CTG CCT CAG GGA TCG ATA AAG TTT CAC TGC AGA AAG TCT CCA CTG CGG TAT GCT
L   P   Q   G   S   I   K   F   H   C   R   K   S   P   L   R   Y   A
```

FIGURE 1E

```
       1846          1855          1864          1873          1882          1891
GAC ATC TGC CCT GAA CCT TCA CCC TAC AGC ATT ACA GGC TTT AAT CAG ATT CTG
D   I   C   P   E   P   S   P   Y   S   I   T   G   F   N   Q   I   L

       1900          1909          1918          1927          1936          1945
CTG GAA AGA CAC AGG CTG ATC CAC GTG ACC TCT TCT GCC TTC ACT GGG CTG GGG
L   E   R   H   R   L   I   H   V   T   S   S   A   F   T   G   L   G

       1954          1963          1972          1981          1990          1999
TGA TCC TTG GTG CCT TTG TTT CCA CAA GGC CTT TTC CTG CCC CCT GCC TTG CCA

       2008          2017          2026          2035          2044          2053
AAG ACA TTT AAT CAG CAC ACA GCT GCC AGA CTA TTC CCA CAG TGC TCC AAA TGC

       2062          2071          2080          2089          2098          2107
ACA TGA ACA ACA GTG ACG GGC TCA GCT TCG ACC CAG AGC CCC GTG CCC AGT GCG

T 3'
```

FIGURE 1F

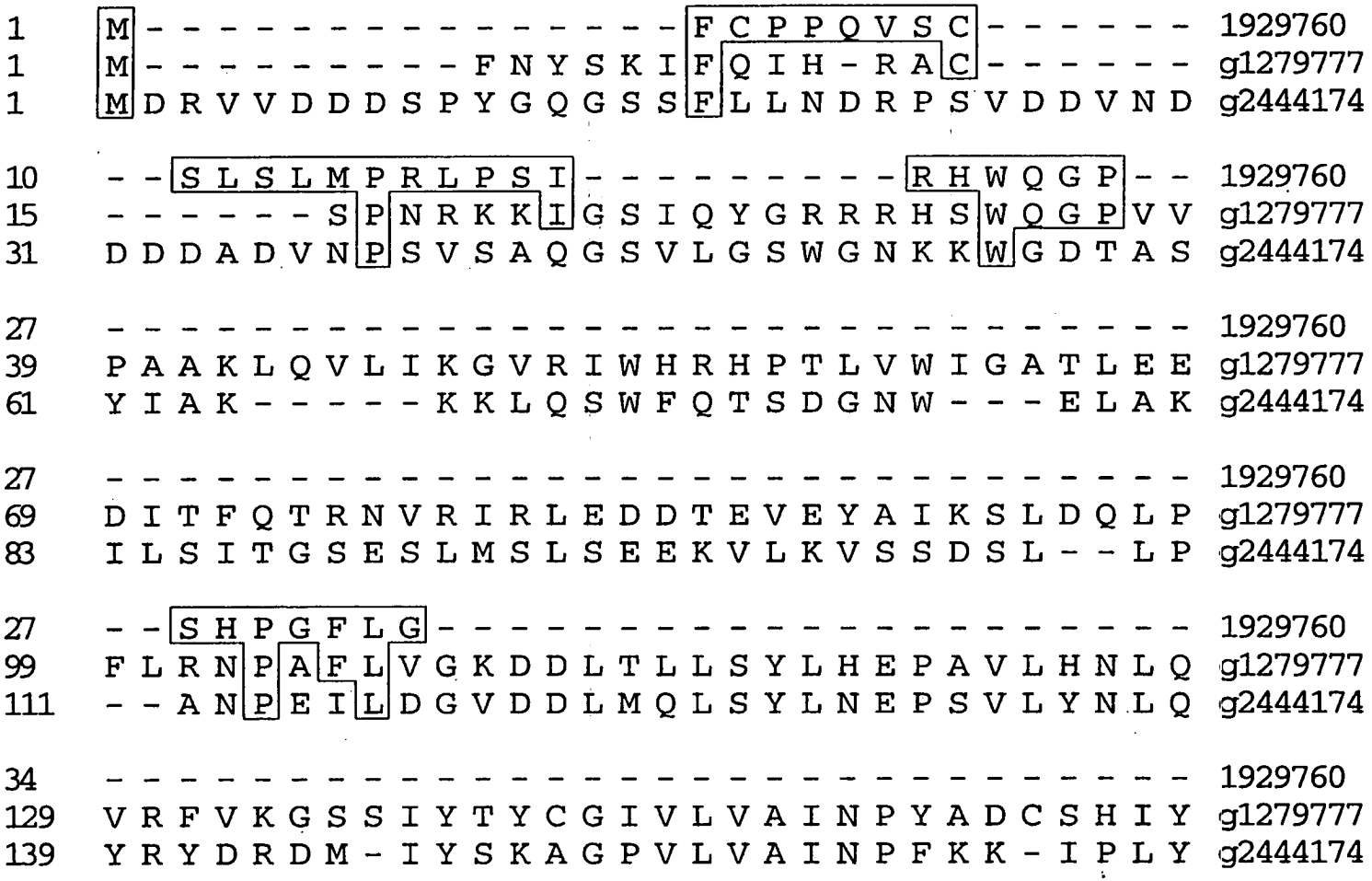
1 M - - - - - - - - - - - - - - - - - - F C P P Q V S C - - - - - - - 1929760
1 M - - - - - - - - - - F N Y S K I F Q I H - R A C - - - - - - - g1279777
1 M D R V V D D D S P Y G Q G S S F L L N D R P S V D D V N D g2444174
10 - - S L S L M P R L P S I - - - - - - - - - - R H W Q G P - - 1929760
15 - - - - - - - S P N R K K I G S I Q Y G R R R H S W Q G P V V g1279777
31 D D D A D V N P S V S A Q G S V L G S W G N K K W G D T A S g2444174
27 - 1929760
39 P A A K L Q V L I K G V R I W H R H P T L V W I G A T L E E g1279777
61 Y I A K - - - - - K K L Q S W F Q T S D G N W - - - E L A K g2444174
27 - 1929760
69 D I T F Q T R N V R I R L E D D T E V E Y A I K S L D Q L P g1279777
83 I L S I T G S E S L M S L S E E K V L K V S S D S L - - L P g2444174
27 - - S H P G F L G - 1929760
99 F L R N P A F L V G K D D L T L L S Y L H E P A V L H N L Q g1279777
111 - - A N P E I L D G V D D L M Q L S Y L N E P S V L Y N L Q g2444174
34 - 1929760
129 V R F V K G S S I Y T Y C G I V L V A I N P Y A D C S H I Y g1279777
139 Y R Y D R D M - I Y S K A G P V L V A I N P F K K - I P L Y g2444174

FIGURE 2A

```
34   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   1929760
159  G E E I I Q V Y R G A G K S A R E M D P H I F A V A E E A H   g1279777
167  G S D Y I E A Y - - - - K R K S I D N P H V Y A I A D T A I   g2444174

34   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   1929760
189  F D M G A F G K S Q S I I V S G E S G A G K T V S A K F V M   g1279777
193  R E M I R D E V N Q S I V I S G E S G A G K T E T P K I A M   g2444174

34   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   1929760
219  R Y L A S V A A S K T R N G G T T S I E A R V L A S N P I M   g1279777
223  Q Y L A A L G G G D A R E S G I L S H N G C R T P R R A - -   g2444174

34   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   1929760
249  E S I G N A K T I R N D N S S R F G K F I Q I N F C E R G R   g1279777
251  E A F G N A K T S R D N N S S R I G K L I E I H F S E T G K   g2444174

34   - - - - - - - - - - - - - - - - - - - - - - - - - - - - P L   1929760
279  R I V G A E M K T Y L L E K S R L V F Q A P G E R N Y H I F   g1279777
281  - I S G A K I Q T F L L E K S R V V Q C T D G E R S Y H S F   g2444174

36   F P I C S - - - - - - - - - - - - - - - - - - - - - - - - -   1929760
309  Y Q L C A A R N H Q V L K D L H L G P C E S Y S Y L T Q G G   g1279777
310  Y Q L C A G A P P S L R E K L N L K S A R E Y K Y F Q Q S T   g2444174
```

FIGURE 2B

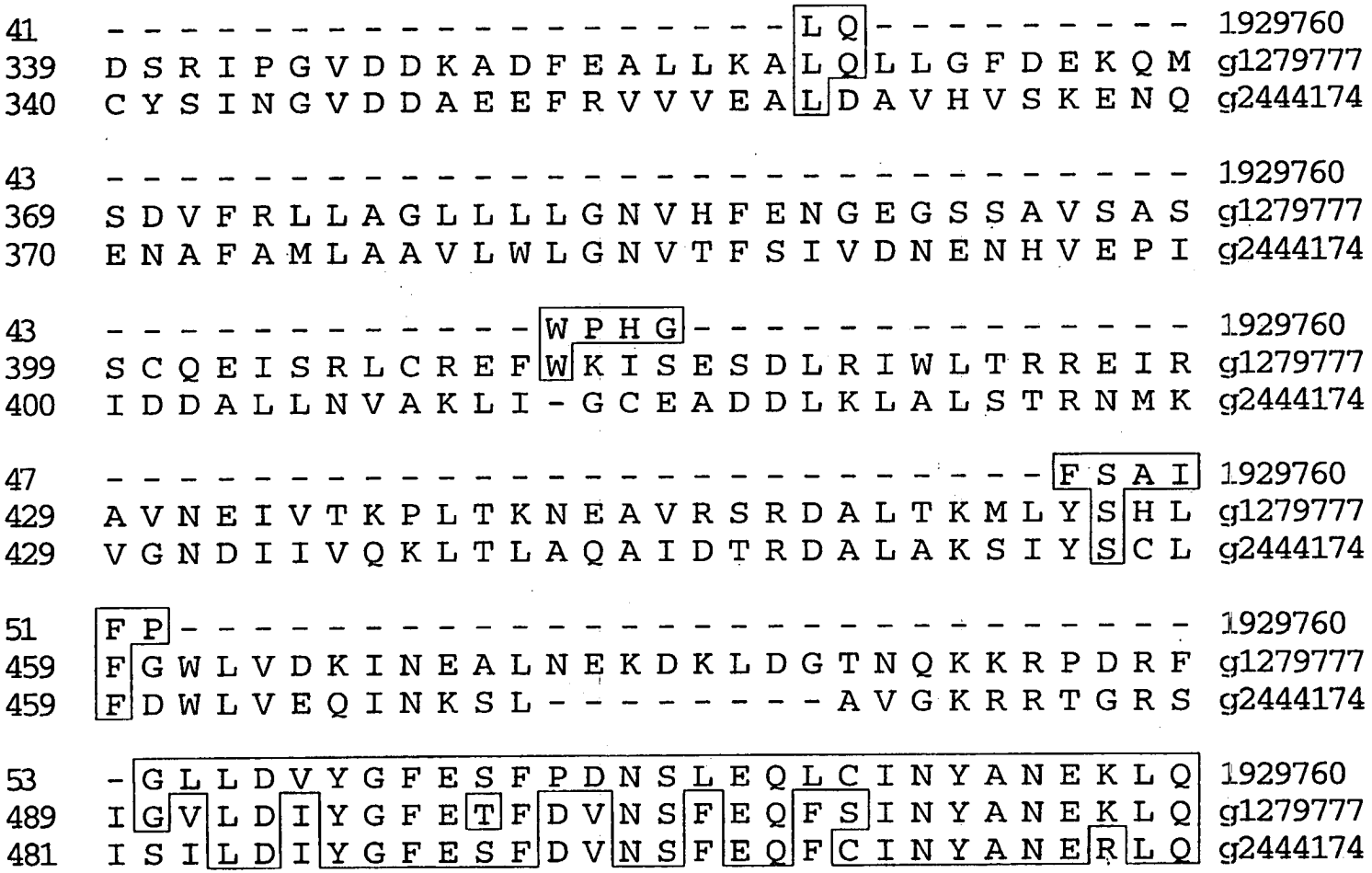
41 - L Q - - - - - - - - - - 1929760
339 D S R I P G V D D K A D F E A L L K A L Q L L G F D E K Q M g1279777
340 C Y S I N G V D D A E E F R V V V E A L D A V H V S K E N Q g2444174
43 - 1929760
369 S D V F R L L A G L L L L G N V H F E N G E G S S A V S A S g1279777
370 E N A F A M L A A V L W L G N V T F S I V D N E N H V E P I g2444174
43 - - - - - - - - - - - - - - - W P H G - - - - - - - - - - - - - - - - 1929760
399 S C Q E I S R L C R E F W K I S E S D L R I W L T R R E I R g1279777
400 I D D A L L N V A K L I - G C E A D D L K L A L S T R N M K g2444174
47 - F S A I 1929760
429 A V N E I V T K P L T K N E A V R S R D A L T K M L Y S H L g1279777
429 V G N D I I V Q K L T L A Q A I D T R D A L A K S I Y S C L g2444174
51 F P - 1929760
459 F G W L V D K I N E A L N E K D K L D G T N Q K K R P D R F g1279777
459 F D W L V E Q I N K S L - - - - - - - - - A V G K R R T G R S g2444174
53 - G L L D V Y G F E S F P D N S L E Q L C I N Y A N E K L Q 1929760
489 I G V L D I Y G F E T F D V N S F E Q F S I N Y A N E K L Q g1279777
481 I S I L D I Y G F E S F D V N S F E Q F C I N Y A N E R L Q g2444174

FIGURE 2C

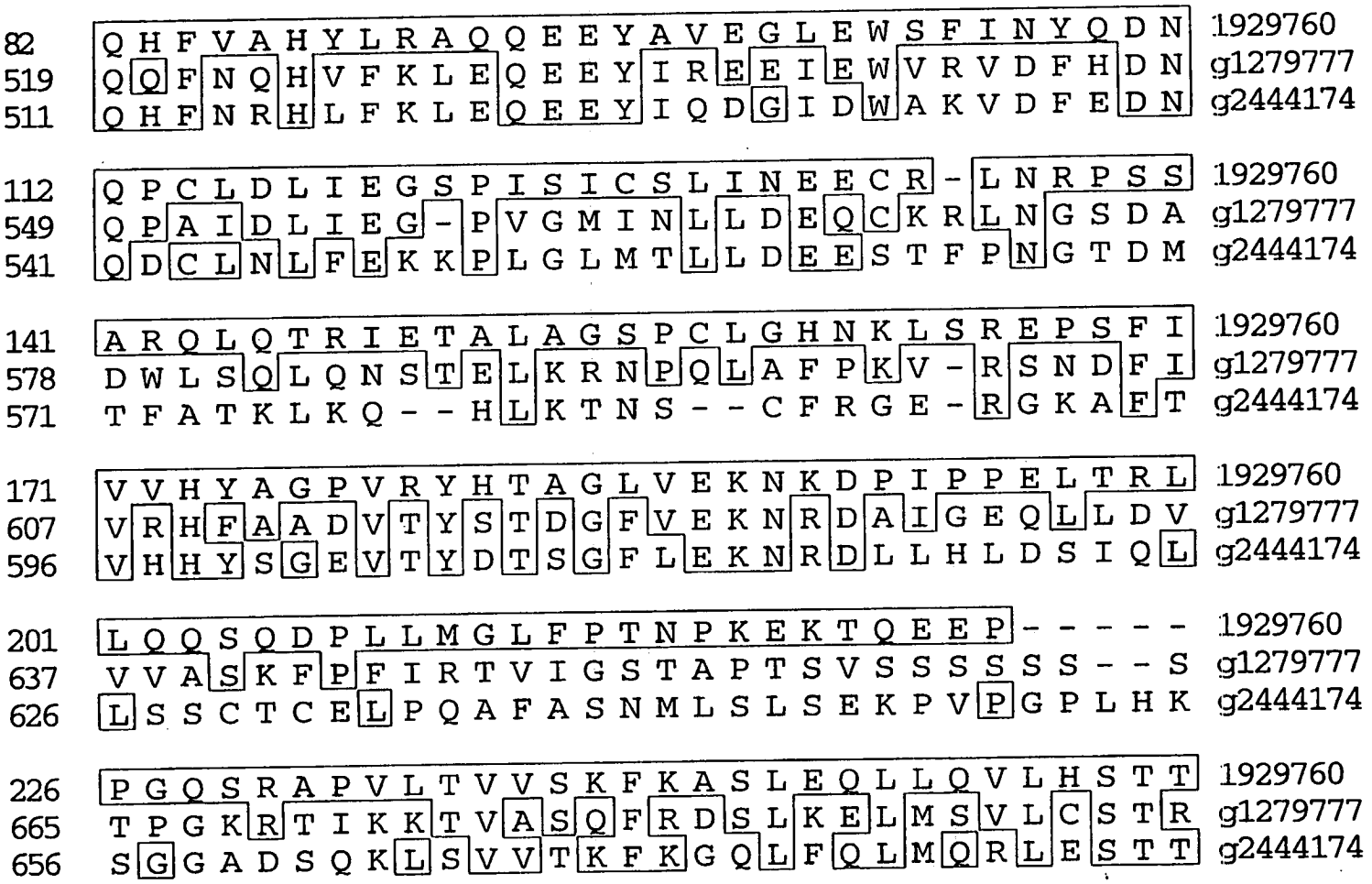
82 Q H F V A H Y L R A Q Q E E Y A V E G L E W S F I N Y Q D N 1929760
519 Q Q F N Q H V F K L E Q E E Y I R E E I E W V R V D F H D N g1279777
511 Q H F N R H L F K L E Q E E Y I Q D G I D W A K V D F E D N g2444174
112 Q P C L D L I E G S P I S I C S L I N E E C R - L N R P S S 1929760
549 Q P A I D L I E G - P V G M I N L L D E Q C K R L N G S D A g1279777
541 Q D C L N L F E K K P L G L M T L L D E E S T F P N G T D M g2444174
141 A R Q L Q T R I E T A L A G S P C L G H N K L S R E P S F I 1929760
578 D W L S Q L Q N S T E L K R N P Q L A F P K V - R S N D F I g1279777
571 T F A T K L K Q - - H L K T N S - - C F R G E - R G K A F T g2444174
171 V V H Y A G P V R Y H T A G L V E K N K D P I P P E L T R L 1929760
607 V R H F A A D V T Y S T D G F V E K N R D A I G E Q L L D V g1279777
596 V H H Y S G E V T Y D T S G F L E K N R D L L H L D S I Q L g2444174
201 L Q Q S Q D P L L M G L F P T N P K E K T Q E E P - - - - - 1929760
637 V V A S K F P F I R T V I G S T A P T S V S S S S S S - - S g1279777
626 L S S C T C E L P Q A F A S N M L S L S E K P V P G P L H K g2444174
226 P G Q S R A P V L T V V S K F K A S L E Q L L Q V L H S T T 1929760
665 T P G K R T I K K T V A S Q F R D S L K E L M S V L C S T R g1279777
656 S G G A D S Q K L S V V T K F K G Q L F Q L M Q R L E S T T g2444174

FIGURE 2D

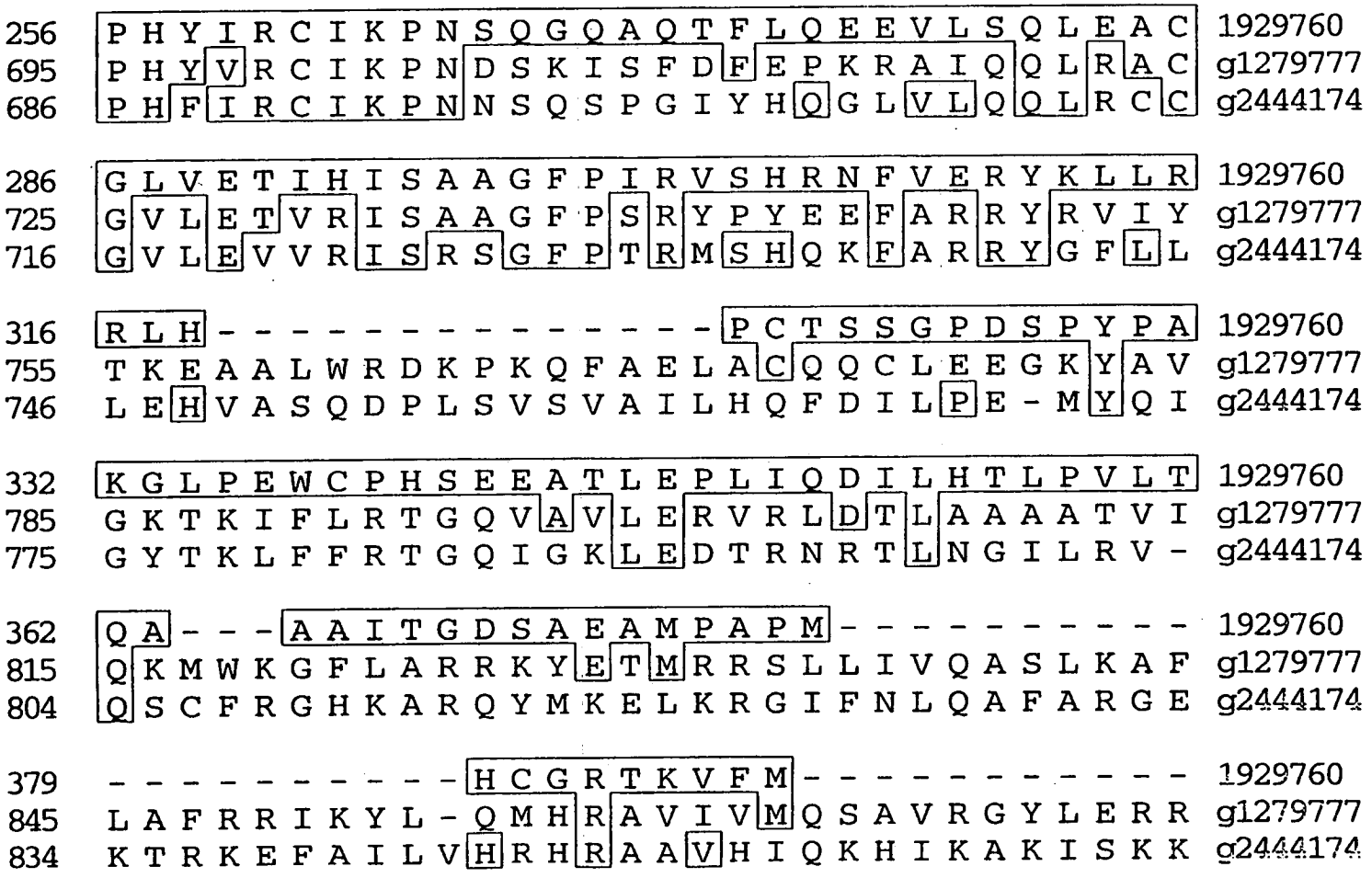
256 P H Y I R C I K P N S Q G Q A Q T F L Q E E V L S Q L E A C 1929760
695 P H Y V R C I K P N D S K I S F D F E P K R A I Q Q L R A C g1279777
686 P H F I R C I K P N N S Q S P G I Y H Q G L V L Q Q L R C C g2444174
286 G L V E T I H I S A A G F P I R V S H R N F V E R Y K L L R 1929760
725 G V L E T V R I S A A G F P S R Y P Y E E F A R R Y R V I Y g1279777
716 G V L E V V R I S R S G F P T R M S H Q K F A R R Y G F L L g2444174
316 R L H - - - - - - - - - - - - - - - - P C T S S G P D S P Y P A 1929760
755 T K E A A L W R D K P K Q F A E L A C Q Q C L E E G K Y A V g1279777
746 L E H V A S Q D P L S V S V A I L H Q F D I L P E - M Y Q I g2444174
332 K G L P E W C P H S E E A T L E P L I Q D I L H T L P V L T 1929760
785 G K T K I F L R T G Q V A V L E R V R L D T L A A A A T V I g1279777
775 G Y T K L F F R T G Q I G K L E D T R N R T L N G I L R V - g2444174
362 Q A - - - A A I T G D S A E A M P A P M - - - - - - - - - - 1929760
815 Q K M W K G F L A R R K Y E T M R R S L L I V Q A S L K A F g1279777
804 Q S C F R G H K A R Q Y M K E L K R G I F N L Q A F A R G E g2444174
379 - - - - - - - - - - H C G R T K V F M - - - - - - - - - - 1929760
845 L A F R R I K Y L - Q M H R A V I V M Q S A V R G Y L E R R g1279777
834 K T R K E F A I L V H R H R A A V H I Q K H I K A K I S K K g2444174

FIGURE 2E

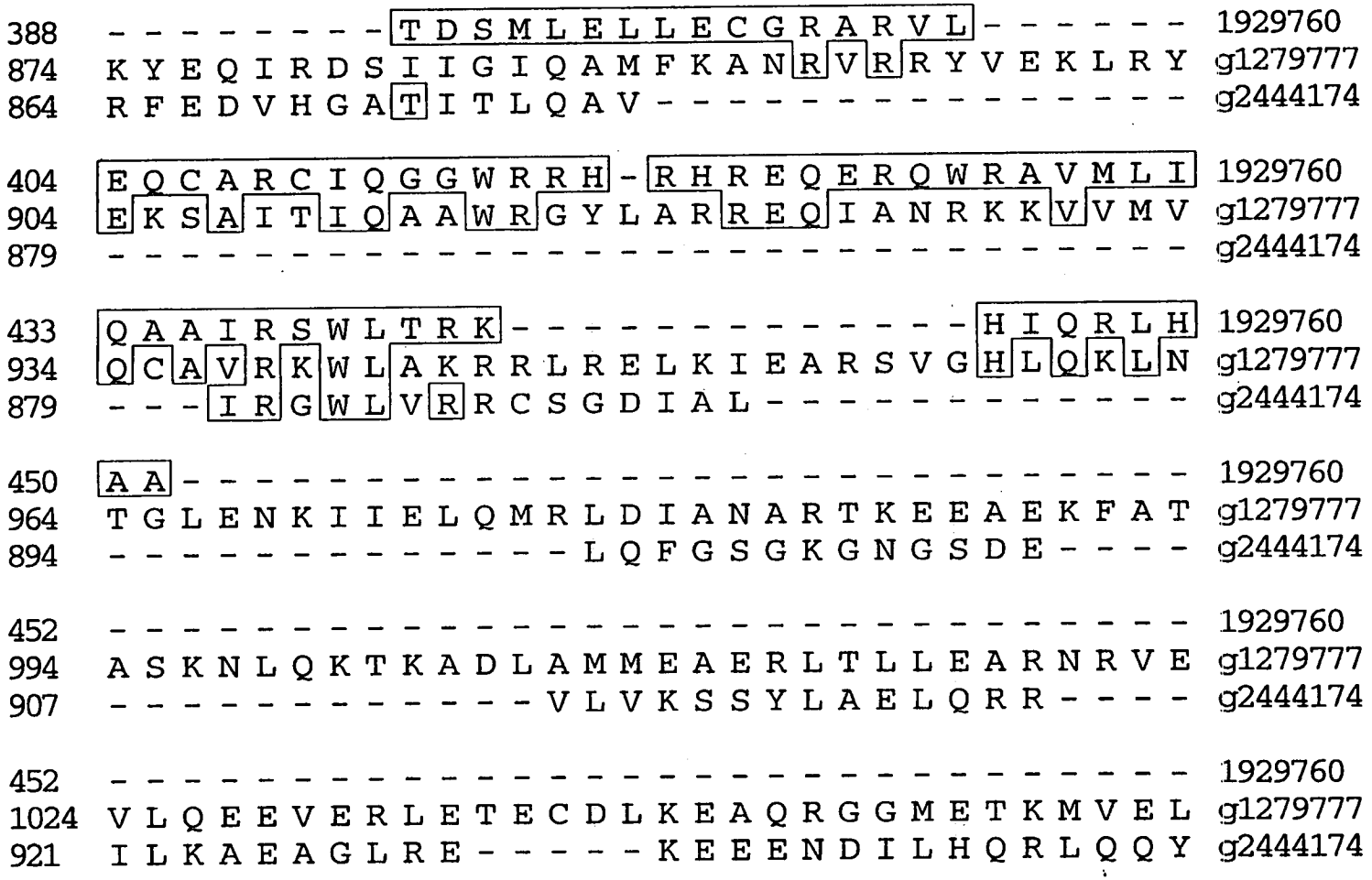
388 - - - - - - - - - T D S M L E L L E C G R A R V L - - - - - - 1929760
874 K Y E Q I R D S I I G I Q A M F K A N R V R R Y V E K L R Y g1279777
864 R F E D V H G A T I T L Q A V - - - - - - - - - - - - - - - - g2444174
404 E Q C A R C I Q G G W R R H - R H R E Q E R Q W R A V M L I 1929760
904 E K S A I T I Q A A W R G Y L A R R E Q I A N R K K V V M V g1279777
879 - g2444174
433 Q A A I R S W L T R K - - - - - - - - - - - - - - H I Q R L H 1929760
934 Q C A V R K W L A K R R L R E L K I E A R S V G H L Q K L N g1279777
879 - - - I R G W L V R R C S G D I A L - - - - - - - - - - - - - g2444174
450 A A - 1929760
964 T G L E N K I I E L Q M R L D I A N A R T K E E A E K F A T g1279777
894 - - - - - - - - - - - - - - - L Q F G S G K G N G S D E - - - - g2444174
452 - 1929760
994 A S K N L Q K T K A D L A M M E A E R L T L L E A R N R V E g1279777
907 - - - - - - - - - - - - - - V L V K S S Y L A E L Q R R - - - - g2444174
452 - 1929760
1024 V L Q E E V E R L E T E C D L K E A Q R G G M E T K M V E L g1279777
921 I L K A E A G L R E - - - - - K E E E N D I L H Q R L Q Q Y g2444174

FIGURE 2F

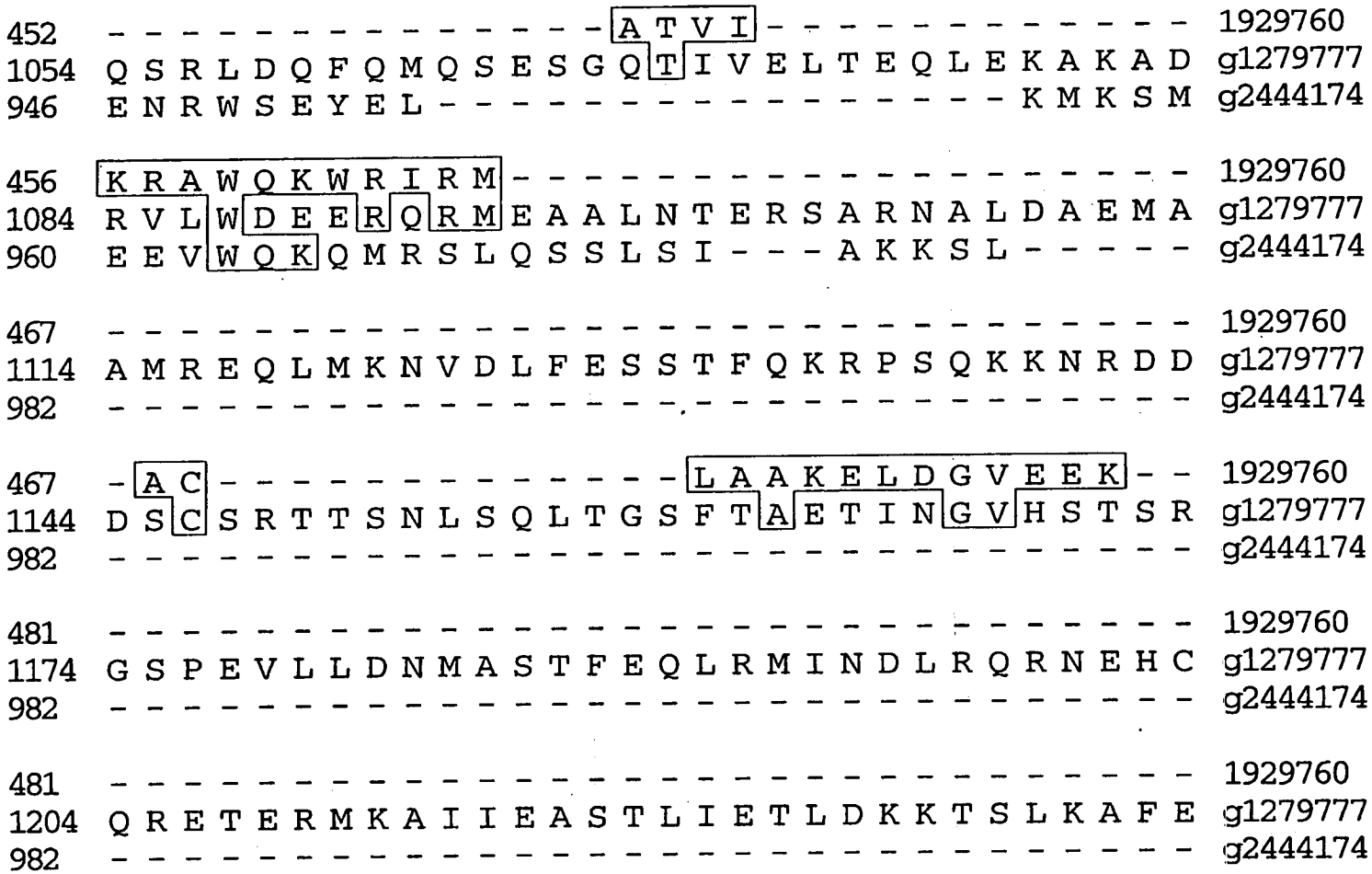
452 - - - - - - - - - - - - - - - - A T V I - - - - - - - - - - - - 1929760
1054 Q S R L D Q F Q M Q S E S G Q T I V E L T E Q L E K A K A D g1279777
946 E N R W S E Y E L - - - - - - - - - - - - - - - - - - K M K S M g2444174
456 K R A W Q K W R I R M - 1929760
1084 R V L W D E E R Q R M E A A L N T E R S A R N A L D A E M A g1279777
960 E E V W Q K Q M R S L Q S S L S I - - - A K K S L - - - - - - g2444174
467 - 1929760
1114 A M R E Q L M K N V D L F E S S T F Q K R P S Q K K N R D D g1279777
982 - g2444174
467 - A C - - - - - - - - - - - - - - - L A A K E L D G V E E K - - 1929760
1144 D S C S R T T S N L S Q L T G S F T A E T I N G V H S T S R g1279777
982 - g2444174
481 - 1929760
1174 G S P E V L L D N M A S T F E Q L R M I N D L R Q R N E H C g1279777
982 - g2444174
481 - 1929760
1204 Q R E T E R M K A I I E A S T L I E T L D K K T S L K A F E g1279777
982 - g2444174

FIGURE 2G

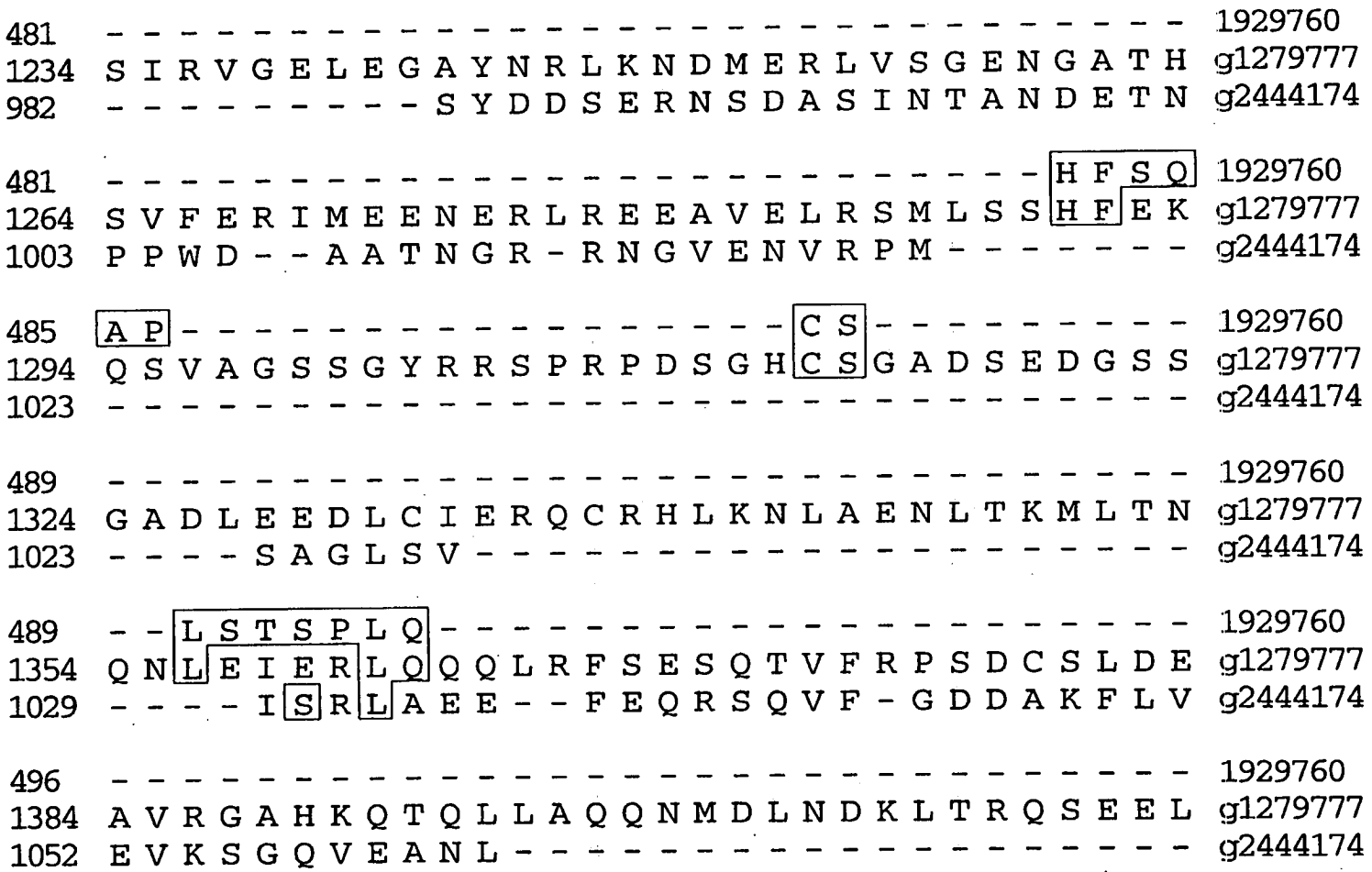
481 - 1929760
1234 S I R V G E L E G A Y N R L K N D M E R L V S G E N G A T H g1279777
982 - - - - - - - - - - S Y D D S E R N S D A S I N T A N D E T N g2444174
481 - H F S Q 1929760
1264 S V F E R I M E E N E R L R E E A V E L R S M L S S H F E K g1279777
1003 P P W D - - A A T N G R - R N G V E N V R P M - - - - - - - - g2444174
485 A P - C S - - - - - - - - - - 1929760
1294 Q S V A G S S G Y R R S P R P D S G H C S G A D S E D G S S g1279777
1023 - g2444174
489 - 1929760
1324 G A D L E E D L C I E R Q C R H L K N L A E N L T K M L T N g1279777
1023 - - - - S A G L S V - g2444174
489 - - L S T S P L Q - 1929760
1354 Q N L E I E R L Q Q Q L R F S E S Q T V F R P S D C S L D E g1279777
1029 - - - - I S R L A E E - - F E Q R S Q V F - G D D A K F L V g2444174
496 - 1929760
1384 A V R G A H K Q T Q L L A Q Q N M D L N D K L T R Q S E E L g1279777
1052 E V K S G Q V E A N L - g2444174

FIGURE 2H

```
496   ------------------------------ 1929760
1414  AEARAQLRGYSGPLGLENASDEEIIRLEAF g1279777
1063  ------------------NPDHELRRLKQ- g2444174

496   ------------------------------ 1929760
1444  EKGSIKHSGFLEVYNVPEFARIIVCELKPT g1279777
1074  ------------------------------ g2444174

496   -TRLL------------------------- 1929760
1474  LARLLTKNLPAYLLVAAFRNHDEKRDETAL g1279777
1074  ------------------------------ g2444174

500   ------------EAIIR--------LWPLG 1929760
1504  TGLFSSVHLVLKDTISRSHDLDLLSLWLVN g1279777
1074  --MFEG------------------------ g2444174

510   L----------------------------- 1929760
1534  LWRLFNLLRQYSGEDSQPEWHVANTETQNS g1279777
1078  -WK-----KDYTA----------------- g2444174

511   ------------------------------ 1929760
1564  YRFKAYDVAPIRDQLKLRIEECYTSLMKKA g1279777
1085  ---------------RLR------------ g2444174
```

FIGURE 2I

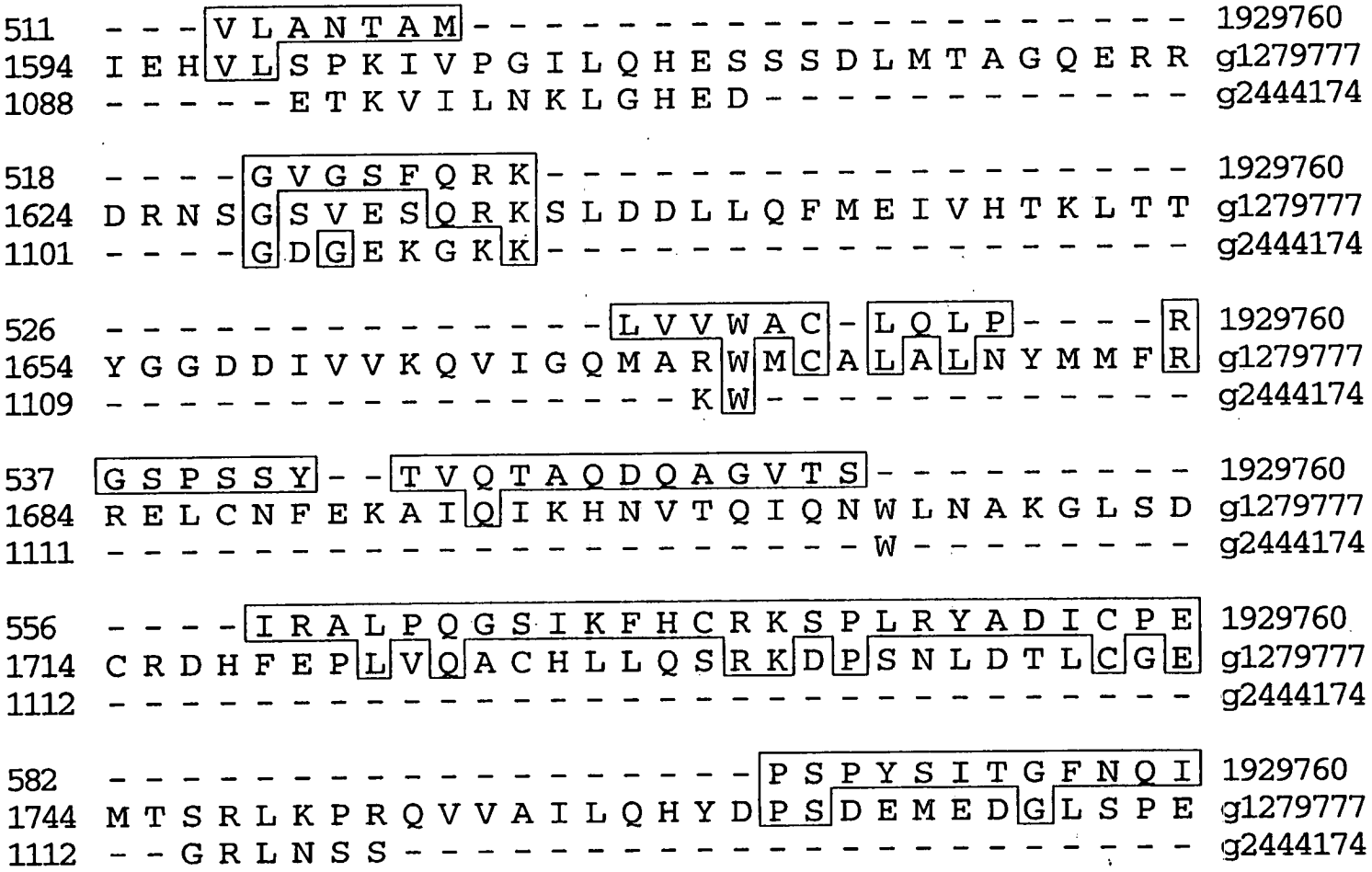
511 - - - V L A N T A M - 1929760
1594 I E H V L S P K I V P G I L Q H E S S S D L M T A G Q E R R g1279777
1088 - - - - - E T K V I L N K L G H E D - - - - - - - - - - - - - g2444174
518 - - - - G V G S F Q R K - - - - - - - - - - - - - - - - - - - 1929760
1624 D R N S G S V E S Q R K S L D D L L Q F M E I V H T K L T T g1279777
1101 - - - - G D G E K G K K - - - - - - - - - - - - - - - - - - - g2444174
526 - - - - - - - - - - - - - - - - L V V W A C - L Q L P - - - - R 1929760
1654 Y G G D D I V V K Q V I G Q M A R W M C A L A L N Y M M F R g1279777
1109 - - - - - - - - - - - - - - - - - - K W - - - - - - - - - - - g2444174
537 G S P S S Y - - T V Q T A Q D Q A G V T S - - - - - - - - - - 1929760
1684 R E L C N F E K A I Q I K H N V T Q I Q N W L N A K G L S D g1279777
1111 - W - - - - - - - - g2444174
556 - - - - I R A L P Q G S I K F H C R K S P L R Y A D I C P E 1929760
1714 C R D H F E P L V Q A C H L L Q S R K D P S N L D T L C G E g1279777
1112 - g2444174
582 - P S P Y S I T G F N Q I 1929760
1744 M T S R L K P R Q V V A I L Q H Y D P S D E M E D G L S P E g1279777
1112 - - G R L N S S - g2444174

FIGURE 2J

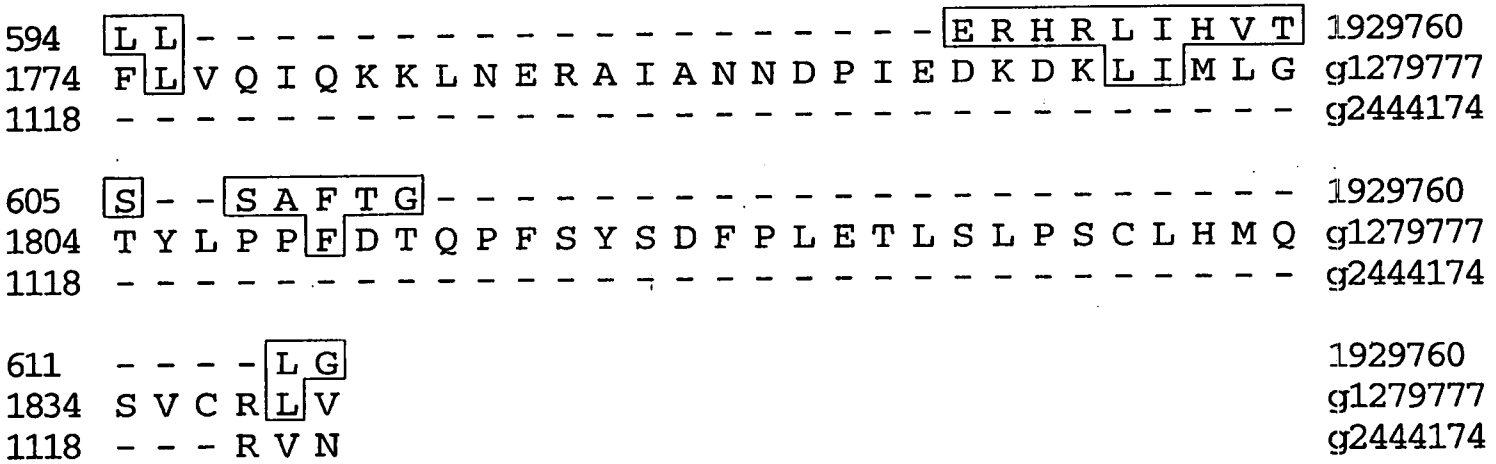
594 L L - E R H R L I H V T 1929760
1774 F L V Q I Q K K L N E R A I A N N D P I E D K D K L I M L G g1279777
1118 - g2444174
605 S - - S A F T G - 1929760
1804 T Y L P P F D T Q P F S Y S D F P L E T L S L P S C L H M Q g1279777
1118 - g2444174
611 - - - - L G 1929760
1834 S V C R L V g1279777
1118 - - - R V N g2444174

FIGURE 2K

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
TANG, Y. Tom
CORLEY, Neil C.
GORGONE, Gina A.
GUEGLER, Karl J.
BAUGHN, Mariah R.

<120> MYOSIN HEAVY CHAIN HOMOLOG

<130> PF-0621 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/187,060; unassigned
<151> 1998-11-05; 1998-11-05

<160> 4

<170> PERL Program

<210> 1
<211> 612
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1929760CD1

<400> 1
Met Phe Cys Pro Pro Gln Val Ser Cys Ser Leu Ser Leu Met Pro
1 5 10 15
Arg Leu Pro Ser Ile Arg His Trp Gln Gly Pro Ser His Pro Gly
20 25 30
Phe Leu Gly Pro Leu Phe Pro Ile Cys Ser Leu Gln Trp Pro His
35 40 45
Gly Phe Ser Ala Ile Phe Pro Gly Leu Leu Asp Val Tyr Gly Phe
50 55 60
Glu Ser Phe Pro Asp Asn Ser Leu Glu Gln Leu Cys Ile Asn Tyr
65 70 75
Ala Asn Glu Lys Leu Gln Gln His Phe Val Ala His Tyr Leu Arg
80 85 90
Ala Gln Gln Glu Glu Tyr Ala Val Glu Gly Leu Glu Trp Ser Phe
95 100 105
Ile Asn Tyr Gln Asp Asn Gln Pro Cys Leu Asp Leu Ile Glu Gly
110 115 120
Ser Pro Ile Ser Ile Cys Ser Leu Ile Asn Glu Glu Cys Arg Leu
125 130 135
Asn Arg Pro Ser Ser Ala Arg Gln Leu Gln Thr Arg Ile Glu Thr
140 145 150
Ala Leu Ala Gly Ser Pro Cys Leu Gly His Asn Lys Leu Ser Arg

```
                155                 160                 165
Glu Pro Ser Phe Ile Val Val His Tyr Ala Gly Pro Val Arg Tyr
                170                 175                 180
His Thr Ala Gly Leu Val Glu Lys Asn Lys Asp Pro Ile Pro Pro
                185                 190                 195
Glu Leu Thr Arg Leu Leu Gln Gln Ser Gln Asp Pro Leu Leu Met
                200                 205                 210
Gly Leu Phe Pro Thr Asn Pro Lys Glu Lys Thr Gln Glu Glu Pro
                215                 220                 225
Pro Gly Gln Ser Arg Ala Pro Val Leu Thr Val Val Ser Lys Phe
                230                 235                 240
Lys Ala Ser Leu Glu Gln Leu Leu Gln Val Leu His Ser Thr Thr
                245                 250                 255
Pro His Tyr Ile Arg Cys Ile Lys Pro Asn Ser Gln Gly Gln Ala
                260                 265                 270
Gln Thr Phe Leu Gln Glu Glu Val Leu Ser Gln Leu Glu Ala Cys
                275                 280                 285
Gly Leu Val Glu Thr Ile His Ile Ser Ala Ala Gly Phe Pro Ile
                290                 295                 300
Arg Val Ser His Arg Asn Phe Val Glu Arg Tyr Lys Leu Leu Arg
                305                 310                 315
Arg Leu His Pro Cys Thr Ser Ser Gly Pro Asp Ser Pro Tyr Pro
                320                 325                 330
Ala Lys Gly Leu Pro Glu Trp Cys Pro His Ser Glu Glu Ala Thr
                335                 340                 345
Leu Glu Pro Leu Ile Gln Asp Ile Leu His Thr Leu Pro Val Leu
                350                 355                 360
Thr Gln Ala Ala Ala Ile Thr Gly Asp Ser Ala Glu Ala Met Pro
                365                 370                 375
Ala Pro Met His Cys Gly Arg Thr Lys Val Phe Met Thr Asp Ser
                380                 385                 390
Met Leu Glu Leu Leu Glu Cys Gly Arg Ala Arg Val Leu Glu Gln
                395                 400                 405
Cys Ala Arg Cys Ile Gln Gly Gly Trp Arg Arg His Arg His Arg
                410                 415                 420
Glu Gln Glu Arg Gln Trp Arg Ala Val Met Leu Ile Gln Ala Ala
                425                 430                 435
Ile Arg Ser Trp Leu Thr Arg Lys His Ile Gln Arg Leu His Ala
                440                 445                 450
Ala Ala Thr Val Ile Lys Arg Ala Trp Gln Lys Trp Arg Ile Arg
                455                 460                 465
Met Ala Cys Leu Ala Ala Lys Glu Leu Asp Gly Val Glu Glu Lys
                470                 475                 480
His Phe Ser Gln Ala Pro Cys Ser Leu Ser Thr Ser Pro Leu Gln
                485                 490                 495
Thr Arg Leu Leu Glu Ala Ile Ile Arg Leu Trp Pro Leu Gly Leu
                500                 505                 510
Val Leu Ala Asn Thr Ala Met Gly Val Gly Ser Phe Gln Arg Lys
                515                 520                 525
Leu Val Val Trp Ala Cys Leu Gln Leu Pro Arg Gly Ser Pro Ser
                530                 535                 540
Ser Tyr Thr Val Gln Thr Ala Gln Asp Gln Ala Gly Val Thr Ser
                545                 550                 555
Ile Arg Ala Leu Pro Gln Gly Ser Ile Lys Phe His Cys Arg Lys
                560                 565                 570
```

Ser Pro Leu Arg Tyr Ala Asp Ile Cys Pro Glu Pro Ser Pro Tyr
575 580 585
Ser Ile Thr Gly Phe Asn Gln Ile Leu Leu Glu Arg His Arg Leu
590 595 600
Ile His Val Thr Ser Ser Ala Phe Thr Gly Leu Gly
605 610

<210> 2
<211> 2109
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1929760CB1

<400> 2
tgatgctctg ggctgtcttc acacttcatt tgggtttcct gcttgctctg agctctacag 60
gggaatgggg tagagatggg agccaccttg ggtggagggt ggggaaggta tgttctgccc 120
accacaggtg tcatgctcac tcagcctgat gcccaggctg ccaagtataa ggcattggca 180
ggggcccagc caccctgggt tccttggtcc cctattcccc atctgctccc tgcagtggcc 240
ccatgggttc tctgccatct tcccaggcct gctggatgtg tatggatttg aatcatttcc 300
tgacaacagt ctggaacagt tgtgcatcaa ctacgccaat gagaagctgc agcagcattt 360
tgtggctcac tacctaaggg cccagcagga ggaatacgca gttgagggcc tggagtggtc 420
attcatcaac taccaggaca accagccctg tttggatctc attgagggaa gccccatcag 480
catctgctcc ctcataaatg aggaatgccg cctcaatcga cccagcagcg cacgccagct 540
ccagacacgc attgagactg ccctggcagg cagcccctgc ctgggccaca ataagctcag 600
ccgggagccc agcttcattg tggtgcatta tgcggggcct gtgcggtacc acacagcagg 660
cctggtggag aagaacaagg accctatccc acctgagctg accaggctcc tgcagcaatc 720
ccaggacccc ctgctcatgg ggctgtttcc tactaacccc aaagagaaga cccaggagga 780
acccccctggc cagagcaggg cccctgtgtt gaccgtggtg tccaagttca aggcctcact 840
ggagcagctt ctgcaggtcc tacacagcac cacgccccac tacattcgct gcatcaagcc 900
caacagccag ggccaggcgc agacctttct ccaagaggag gtcctgagcc agctggaggc 960
ctgtggcctc gtggagacca tccatatcag tgctgctggc ttccccatcc gggtctctca 1020
ccgaaacttt gtagaacgat acaagttact aagaaggctt catccttgca catcctctgg 1080
ccccgacagc ccatatcctg ccaaagggct ccctgaatgg tgtccacaca gcgaggaagc 1140
cacgcttgaa cctctcatcc aggacattct ccacactctg ccggtcctaa ctcaggcagc 1200
agccataact ggtgactcgg ctgaggccat gccagccccc atgcactgtg gcaggaccaa 1260
ggtgttcatg actgactcta tgctggagct tctggaatgt gggcgtgccc gggtgctgga 1320
gcagtgtgcc cgctgcatcc agggtggctg gaggcgacac cggcaccgag agcaggagcg 1380
gcagtggcgg gccgtcatgc tcatccaggc agccattcgt tcctggttaa ctcggaaaca 1440
catccagagg ctgcatgcag ctgccacagt catcaagcgt gcatggcaga agtggagaat 1500
cagaatggcc tgccttgctg ctaaagagct ggatggtgtg gaagaaaaac acttctctca 1560
agctccctgt tccctgagca cctcgccgct gcagaccagg ctcctggagg caataatccg 1620
cctctggccc ctgggactgg tcctggccaa tacggctatg ggtgtaggca gctttcagag 1680
gaaattagtg gtctgggctt gcctccagct ccccaggggc agccccagta gctacactgt 1740
ccagacagca caagaccagg ctggtgtcac gtccatccga gcgctgcctc agggatcgat 1800
aaagtttcac tgcagaaagt ctccactgcg gtatgctgac atctgccctg aaccttcacc 1860
ctacagcatt acaggcttta atcagattct gctggaaaga cacaggctga tccacgtgac 1920
ctcttctgcc ttcactgggc tggggtgatc cttggtgcct ttgtttccac aaggcctttt 1980
cctgcccccc gccttgccaa agacatttaa tcagcacaca gctgccagac tattcccaca 2040
gtgctccaaa tgcacatgaa caacagtgac ggctccagcc ttcgacccag agccccgtgc 2100
ccagtgcgt 2109

<210> 3
<211> 1839
<212> PRT
<213> Caenorhabditis elegans

<300>
<308> GenBank ID No: g1279777

<400> 3
Met Phe Asn Tyr Ser Lys Ile Phe Gln Ile His Arg Ala Cys Ser
1 5 10 15
Pro Asn Arg Lys Lys Ile Gly Ser Ile Gln Tyr Gly Arg Arg Arg
20 25 30
His Ser Trp Gln Gly Pro Val Val Pro Ala Ala Lys Leu Gln Val
35 40 45
Leu Ile Lys Gly Val Arg Ile Trp His Arg His Pro Thr Leu Val
50 55 60
Trp Ile Gly Ala Thr Leu Glu Glu Asp Ile Thr Phe Gln Thr Arg
65 70 75
Asn Val Arg Ile Arg Leu Glu Asp Asp Thr Glu Val Glu Tyr Ala
80 85 90
Ile Lys Ser Leu Asp Gln Leu Pro Phe Leu Arg Asn Pro Ala Phe
95 100 105
Leu Val Gly Lys Asp Asp Leu Thr Leu Leu Ser Tyr Leu His Glu
110 115 120
Pro Ala Val Leu His Asn Leu Gln Val Arg Phe Val Lys Gly Ser
125 130 135
Ser Ile Tyr Thr Tyr Cys Gly Ile Val Leu Val Ala Ile Asn Pro
140 145 150
Tyr Ala Asp Cys Ser His Ile Tyr Gly Glu Glu Ile Ile Gln Val
155 160 165
Tyr Arg Gly Ala Gly Lys Ser Ala Arg Glu Met Asp Pro His Ile
170 175 180
Phe Ala Val Ala Glu Glu Ala His Phe Asp Met Gly Ala Phe Gly
185 190 195
Lys Ser Gln Ser Ile Ile Val Ser Gly Glu Ser Gly Ala Gly Lys
200 205 210
Thr Val Ser Ala Lys Phe Val Met Arg Tyr Leu Ala Ser Val Ala
215 220 225
Ala Ser Lys Thr Arg Asn Gly Gly Thr Thr Ser Ile Glu Ala Arg
230 235 240
Val Leu Ala Ser Asn Pro Ile Met Glu Ser Ile Gly Asn Ala Lys
245 250 255
Thr Ile Arg Asn Asp Asn Ser Ser Arg Phe Gly Lys Phe Ile Gln
260 265 270
Ile Asn Phe Cys Glu Arg Gly Arg Arg Ile Val Gly Ala Glu Met
275 280 285
Lys Thr Tyr Leu Leu Glu Lys Ser Arg Leu Val Phe Gln Ala Pro
290 295 300
Gly Glu Arg Asn Tyr His Ile Phe Tyr Gln Leu Cys Ala Ala Arg
305 310 315
Asn His Gln Val Leu Lys Asp Leu His Leu Gly Pro Cys Glu Ser
320 325 330
Tyr Ser Tyr Leu Thr Gln Gly Gly Asp Ser Arg Ile Pro Gly Val

```
                335                 340                 345
Asp Asp Lys Ala Asp Phe Glu Ala Leu Leu Lys Ala Leu Gln Leu
                350                 355                 360
Leu Gly Phe Asp Glu Lys Gln Met Ser Asp Val Phe Arg Leu Leu
                365                 370                 375
Ala Gly Leu Leu Leu Leu Gly Asn Val His Phe Glu Asn Gly Glu
                380                 385                 390
Gly Ser Ser Ala Val Ser Ala Ser Ser Cys Gln Glu Ile Ser Arg
                395                 400                 405
Leu Cys Arg Glu Phe Trp Lys Ile Ser Glu Ser Asp Leu Arg Ile
                410                 415                 420
Trp Leu Thr Arg Arg Glu Ile Arg Ala Val Asn Glu Ile Val Thr
                425                 430                 435
Lys Pro Leu Thr Lys Asn Glu Ala Val Arg Ser Arg Asp Ala Leu
                440                 445                 450
Thr Lys Met Leu Tyr Ser His Leu Phe Gly Trp Leu Val Asp Lys
                455                 460                 465
Ile Asn Glu Ala Leu Asn Glu Lys Asp Lys Leu Asp Gly Thr Asn
                470                 475                 480
Gln Lys Lys Arg Pro Asp Arg Phe Ile Gly Val Leu Asp Ile Tyr
                485                 490                 495
Gly Phe Glu Thr Phe Asp Val Asn Ser Phe Glu Gln Phe Ser Ile
                500                 505                 510
Asn Tyr Ala Asn Glu Lys Leu Gln Gln Gln Phe Asn Gln His Val
                515                 520                 525
Phe Lys Leu Glu Gln Glu Glu Tyr Ile Arg Glu Glu Ile Glu Trp
                530                 535                 540
Val Arg Val Asp Phe His Asp Asn Gln Pro Ala Ile Asp Leu Ile
                545                 550                 555
Glu Gly Pro Val Gly Met Ile Asn Leu Leu Asp Glu Gln Cys Lys
                560                 565                 570
Arg Leu Asn Gly Ser Asp Ala Asp Trp Leu Ser Gln Leu Gln Asn
                575                 580                 585
Ser Thr Glu Leu Lys Arg Asn Pro Gln Leu Ala Phe Pro Lys Val
                590                 595                 600
Arg Ser Asn Asp Phe Ile Val Arg His Phe Ala Ala Asp Val Thr
                605                 610                 615
Tyr Ser Thr Asp Gly Phe Val Glu Lys Asn Arg Asp Ala Ile Gly
                620                 625                 630
Glu Gln Leu Leu Asp Val Val Val Ala Ser Lys Phe Pro Phe Ile
                635                 640                 645
Arg Thr Val Ile Gly Ser Thr Ala Pro Thr Ser Val Ser Ser Ser
                650                 655                 660
Ser Ser Ser Ser Thr Pro Gly Lys Arg Thr Ile Lys Lys Thr Val
                665                 670                 675
Ala Ser Gln Phe Arg Asp Ser Leu Lys Glu Leu Met Ser Val Leu
                680                 685                 690
Cys Ser Thr Arg Pro His Tyr Val Arg Cys Ile Lys Pro Asn Asp
                695                 700                 705
Ser Lys Ile Ser Phe Asp Phe Glu Pro Lys Arg Ala Ile Gln Gln
                710                 715                 720
Leu Arg Ala Cys Gly Val Leu Glu Thr Val Arg Ile Ser Ala Ala
                725                 730                 735
Gly Phe Pro Ser Arg Tyr Pro Tyr Glu Glu Phe Ala Arg Arg Tyr
                740                 745                 750
```

```
Arg Val Ile Tyr Thr Lys Glu Ala Ala Leu Trp Arg Asp Lys Pro
                755                 760                 765
Lys Gln Phe Ala Glu Leu Ala Cys Gln Gln Cys Leu Glu Glu Gly
                770                 775                 780
Lys Tyr Ala Val Gly Lys Thr Lys Ile Phe Leu Arg Thr Gly Gln
                785                 790                 795
Val Ala Val Leu Glu Arg Val Arg Leu Asp Thr Leu Ala Ala Ala
                800                 805                 810
Ala Thr Val Ile Gln Lys Met Trp Lys Gly Phe Leu Ala Arg Arg
                815                 820                 825
Lys Tyr Glu Thr Met Arg Arg Ser Leu Leu Ile Val Gln Ala Ser
                830                 835                 840
Leu Lys Ala Phe Leu Ala Phe Arg Arg Ile Lys Tyr Leu Gln Met
                845                 850                 855
His Arg Ala Val Ile Val Met Gln Ser Ala Val Arg Gly Tyr Leu
                860                 865                 870
Glu Arg Arg Lys Tyr Glu Gln Ile Arg Asp Ser Ile Ile Gly Ile
                875                 880                 885
Gln Ala Met Phe Lys Ala Asn Arg Val Arg Arg Tyr Val Glu Lys
                890                 895                 900
Leu Arg Tyr Glu Lys Ser Ala Ile Thr Ile Gln Ala Ala Trp Arg
                905                 910                 915
Gly Tyr Leu Ala Arg Arg Glu Gln Ile Ala Asn Arg Lys Lys Val
                920                 925                 930
Val Met Val Gln Cys Ala Val Arg Lys Trp Leu Ala Lys Arg Arg
                935                 940                 945
Leu Arg Glu Leu Lys Ile Glu Ala Arg Ser Val Gly His Leu Gln
                950                 955                 960
Lys Leu Asn Thr Gly Leu Glu Asn Lys Ile Ile Glu Leu Gln Met
                965                 970                 975
Arg Leu Asp Ile Ala Asn Ala Arg Thr Lys Glu Glu Ala Glu Lys
                980                 985                 990
Phe Ala Thr Ala Ser Lys Asn Leu Gln Lys Thr Lys Ala Asp Leu
                995                1000                1005
Ala Met Met Glu Ala Glu Arg Leu Thr Leu Leu Glu Ala Arg Asn
               1010                1015                1020
Arg Val Glu Val Leu Gln Glu Glu Val Glu Arg Leu Glu Thr Glu
               1025                1030                1035
Cys Asp Leu Lys Glu Ala Gln Arg Gly Gly Met Glu Thr Lys Met
               1040                1045                1050
Val Glu Leu Gln Ser Arg Leu Asp Gln Phe Gln Met Gln Ser Glu
               1055                1060                1065
Ser Gly Gln Thr Ile Val Glu Leu Thr Glu Gln Leu Glu Lys Ala
               1070                1075                1080
Lys Ala Asp Arg Val Leu Trp Asp Glu Glu Arg Gln Arg Met Glu
               1085                1090                1095
Ala Ala Leu Asn Thr Glu Arg Ser Ala Arg Asn Ala Leu Asp Ala
               1100                1105                1110
Glu Met Ala Ala Met Arg Glu Gln Leu Met Lys Asn Val Asp Leu
               1115                1120                1125
Phe Glu Ser Ser Thr Phe Gln Lys Arg Pro Ser Gln Lys Lys Asn
               1130                1135                1140
Arg Asp Asp Asp Ser Cys Ser Arg Thr Thr Ser Asn Leu Ser Gln
               1145                1150                1155
Leu Thr Gly Ser Phe Thr Ala Glu Thr Ile Asn Gly Val His Ser
```

```
                1160                1165                1170
Thr Ser Arg Gly Ser Pro Glu Val Leu Leu Asp Asn Met Ala Ser
                1175                1180                1185
Thr Phe Glu Gln Leu Arg Met Ile Asn Asp Leu Arg Gln Arg Asn
                1190                1195                1200
Glu His Cys Gln Arg Glu Thr Glu Arg Met Lys Ala Ile Ile Glu
                1205                1210                1215
Ala Ser Thr Leu Ile Glu Thr Leu Asp Lys Lys Thr Ser Leu Lys
                1220                1225                1230
Ala Phe Glu Ser Ile Arg Val Gly Glu Leu Glu Gly Ala Tyr Asn
                1235                1240                1245
Arg Leu Lys Asn Asp Met Glu Arg Leu Val Ser Gly Glu Asn Gly
                1250                1255                1260
Ala Thr His Ser Val Phe Glu Arg Ile Met Glu Glu Asn Glu Arg
                1265                1270                1275
Leu Arg Glu Glu Ala Val Glu Leu Arg Ser Met Leu Ser Ser His
                1280                1285                1290
Phe Glu Lys Gln Ser Val Ala Gly Ser Ser Gly Tyr Arg Arg Ser
                1295                1300                1305
Pro Arg Pro Asp Ser Gly His Cys Ser Gly Ala Asp Ser Glu Asp
                1310                1315                1320
Gly Ser Ser Gly Ala Asp Leu Glu Glu Asp Leu Cys Ile Glu Arg
                1325                1330                1335
Gln Cys Arg His Leu Lys Asn Leu Ala Glu Asn Leu Thr Lys Met
                1340                1345                1350
Leu Thr Asn Gln Asn Leu Glu Ile Glu Arg Leu Gln Gln Gln Leu
                1355                1360                1365
Arg Phe Ser Glu Ser Gln Thr Val Phe Arg Pro Ser Asp Cys Ser
                1370                1375                1380
Leu Asp Glu Ala Val Arg Gly Ala His Lys Gln Thr Gln Leu Leu
                1385                1390                1395
Ala Gln Gln Asn Met Asp Leu Asn Asp Lys Leu Thr Arg Gln Ser
                1400                1405                1410
Glu Glu Leu Ala Glu Ala Arg Ala Gln Leu Arg Gly Tyr Ser Gly
                1415                1420                1425
Pro Leu Gly Leu Glu Asn Ala Ser Asp Glu Glu Ile Ile Arg Leu
                1430                1435                1440
Glu Ala Phe Glu Lys Gly Ser Ile Lys His Ser Gly Phe Leu Glu
                1445                1450                1455
Val Tyr Asn Val Pro Glu Phe Ala Arg Ile Ile Val Cys Glu Leu
                1460                1465                1470
Lys Pro Thr Leu Ala Arg Leu Leu Thr Lys Asn Leu Pro Ala Tyr
                1475                1480                1485
Leu Leu Val Ala Ala Phe Arg Asn His Asp Glu Lys Arg Asp Glu
                1490                1495                1500
Thr Ala Leu Thr Gly Leu Phe Ser Ser Val His Leu Val Leu Lys
                1505                1510                1515
Asp Thr Ile Ser Arg Ser His Asp Leu Asp Leu Leu Ser Leu Trp
                1520                1525                1530
Leu Val Asn Leu Trp Arg Leu Phe Asn Leu Leu Arg Gln Tyr Ser
                1535                1540                1545
Gly Glu Asp Ser Gln Pro Glu Trp His Val Ala Asn Thr Glu Thr
                1550                1555                1560
Gln Asn Ser Tyr Arg Phe Lys Ala Tyr Asp Val Ala Pro Ile Arg
                1565                1570                1575
```

```
Asp Gln Leu Lys Leu Arg Ile Glu Glu Cys Tyr Thr Ser Leu Met
                1580                1585                1590
Lys Lys Ala Ile Glu His Val Leu Ser Pro Lys Ile Val Pro Gly
                1595                1600                1605
Ile Leu Gln His Glu Ser Ser Ser Asp Leu Met Thr Ala Gly Gln
                1610                1615                1620
Glu Arg Arg Asp Arg Asn Ser Gly Ser Val Glu Ser Gln Arg Lys
                1625                1630                1635
Ser Leu Asp Asp Leu Leu Gln Phe Met Glu Ile Val His Thr Lys
                1640                1645                1650
Leu Thr Thr Tyr Gly Gly Asp Asp Ile Val Val Lys Gln Val Ile
                1655                1660                1665
Gly Gln Met Ala Arg Trp Met Cys Ala Leu Ala Leu Asn Tyr Met
                1670                1675                1680
Met Phe Arg Arg Glu Leu Cys Asn Phe Glu Lys Ala Ile Gln Ile
                1685                1690                1695
Lys His Asn Val Thr Gln Ile Gln Asn Trp Leu Asn Ala Lys Gly
                1700                1705                1710
Leu Ser Asp Cys Arg Asp His Phe Glu Pro Leu Val Gln Ala Cys
                1715                1720                1725
His Leu Leu Gln Ser Arg Lys Asp Pro Ser Asn Leu Asp Thr Leu
                1730                1735                1740
Cys Gly Glu Met Thr Ser Arg Leu Lys Pro Arg Gln Val Val Ala
                1745                1750                1755
Ile Leu Gln His Tyr Asp Pro Ser Asp Glu Met Glu Asp Gly Leu
                1760                1765                1770
Ser Pro Glu Phe Leu Val Gln Ile Gln Lys Lys Leu Asn Glu Arg
                1775                1780                1785
Ala Ile Ala Asn Asn Asp Pro Ile Glu Asp Lys Asp Lys Leu Ile
                1790                1795                1800
Met Leu Gly Thr Tyr Leu Pro Pro Phe Asp Thr Gln Pro Phe Ser
                1805                1810                1815
Tyr Ser Asp Phe Pro Leu Glu Thr Leu Ser Leu Pro Ser Cys Leu
                1820                1825                1830
His Met Gln Ser Val Cys Arg Leu Val
                1835
```

<210> 4
<211> 1120
<212> PRT
<213> Helianthus annuus

<300>
<308> GenBank ID No: g2444174

<400> 4

```
Met Asp Arg Val Val Asp Asp Asp Ser Pro Tyr Gly Gln Gly Ser
  1               5                  10                  15
Ser Phe Leu Leu Asn Asp Arg Pro Ser Val Asp Asp Val Asn Asp
                20                  25                  30
Asp Asp Asp Ala Asp Val Asn Pro Ser Val Ser Ala Gln Gly Ser
                35                  40                  45
Val Leu Gly Ser Trp Gly Asn Lys Lys Trp Gly Asp Thr Ala Ser
```

```
                50                  55                  60
Tyr Ile Ala Lys Lys Lys Leu Gln Ser Trp Phe Gln Thr Ser Asp
                65                  70                  75
Gly Asn Trp Glu Leu Ala Lys Ile Leu Ser Ile Thr Gly Ser Glu
                80                  85                  90
Ser Leu Met Ser Leu Ser Glu Glu Lys Val Leu Lys Val Ser Ser
                95                 100                 105
Asp Ser Leu Leu Pro Ala Asn Pro Glu Ile Leu Asp Gly Val Asp
               110                 115                 120
Asp Leu Met Gln Leu Ser Tyr Leu Asn Glu Pro Ser Val Leu Tyr
               125                 130                 135
Asn Leu Gln Tyr Arg Tyr Asp Arg Asp Met Ile Tyr Ser Lys Ala
               140                 145                 150
Gly Pro Val Leu Val Ala Ile Asn Pro Phe Lys Lys Ile Pro Leu
               155                 160                 165
Tyr Gly Ser Asp Tyr Ile Glu Ala Tyr Lys Arg Lys Ser Ile Asp
               170                 175                 180
Asn Pro His Val Tyr Ala Ile Ala Asp Thr Ala Ile Arg Glu Met
               185                 190                 195
Ile Arg Asp Glu Val Asn Gln Ser Ile Val Ile Ser Gly Glu Ser
               200                 205                 210
Gly Ala Gly Lys Thr Glu Thr Pro Lys Ile Ala Met Gln Tyr Leu
               215                 220                 225
Ala Ala Leu Gly Gly Gly Asp Ala Arg Glu Ser Gly Ile Leu Ser
               230                 235                 240
His Asn Gly Cys Arg Thr Pro Arg Arg Ala Glu Ala Phe Gly Asn
               245                 250                 255
Ala Lys Thr Ser Arg Asp Asn Asn Ser Ser Arg Ile Gly Lys Leu
               260                 265                 270
Ile Glu Ile His Phe Ser Glu Thr Gly Lys Ile Ser Gly Ala Lys
               275                 280                 285
Ile Gln Thr Phe Leu Leu Glu Lys Ser Arg Val Val Gln Cys Thr
               290                 295                 300
Asp Gly Glu Arg Ser Tyr His Ser Phe Tyr Gln Leu Cys Ala Gly
               305                 310                 315
Ala Pro Pro Ser Leu Arg Glu Lys Leu Asn Leu Lys Ser Ala Arg
               320                 325                 330
Glu Tyr Lys Tyr Phe Gln Gln Ser Thr Cys Tyr Ser Ile Asn Gly
               335                 340                 345
Val Asp Asp Ala Glu Glu Phe Arg Val Val Val Glu Ala Leu Asp
               350                 355                 360
Ala Val His Val Ser Lys Glu Asn Gln Glu Asn Ala Phe Ala Met
               365                 370                 375
Leu Ala Ala Val Leu Trp Leu Gly Asn Val Thr Phe Ser Ile Val
               380                 385                 390
Asp Asn Glu Asn His Val Glu Pro Ile Ile Asp Asp Ala Leu Leu
               395                 400                 405
Asn Val Ala Lys Leu Ile Gly Cys Glu Ala Asp Asp Leu Lys Leu
               410                 415                 420
Ala Leu Ser Thr Arg Asn Met Lys Val Gly Asn Asp Ile Ile Val
               425                 430                 435
Gln Lys Leu Thr Leu Ala Gln Ala Ile Asp Thr Arg Asp Ala Leu
               440                 445                 450
Ala Lys Ser Ile Tyr Ser Cys Leu Phe Asp Trp Leu Val Glu Gln
               455                 460                 465
```

```
Ile Asn Lys Ser Leu Ala Val Gly Lys Arg Arg Thr Gly Arg Ser
                470                 475                 480
Ile Ser Ile Leu Asp Ile Tyr Gly Phe Glu Ser Phe Asp Val Asn
                485                 490                 495
Ser Phe Glu Gln Phe Cys Ile Asn Tyr Ala Asn Glu Arg Leu Gln
                500                 505                 510
Gln His Phe Asn Arg His Leu Phe Lys Leu Glu Gln Glu Glu Tyr
                515                 520                 525
Ile Gln Asp Gly Ile Asp Trp Ala Lys Val Asp Phe Glu Asp Asn
                530                 535                 540
Gln Asp Cys Leu Asn Leu Phe Glu Lys Lys Pro Leu Gly Leu Met
                545                 550                 555
Thr Leu Leu Asp Glu Glu Ser Thr Phe Pro Asn Gly Thr Asp Met
                560                 565                 570
Thr Phe Ala Thr Lys Leu Lys Gln His Leu Lys Thr Asn Ser Cys
                575                 580                 585
Phe Arg Gly Glu Arg Gly Lys Ala Phe Thr Val His His Tyr Ser
                590                 595                 600
Gly Glu Val Thr Tyr Asp Thr Ser Gly Phe Leu Glu Lys Asn Arg
                605                 610                 615
Asp Leu Leu His Leu Asp Ser Ile Gln Leu Leu Ser Ser Cys Thr
                620                 625                 630
Cys Glu Leu Pro Gln Ala Phe Ala Ser Asn Met Leu Ser Leu Ser
                635                 640                 645
Glu Lys Pro Val Pro Gly Pro Leu His Lys Ser Gly Gly Ala Asp
                650                 655                 660
Ser Gln Lys Leu Ser Val Val Thr Lys Phe Lys Gly Gln Leu Phe
                665                 670                 675
Gln Leu Met Gln Arg Leu Glu Ser Thr Thr Pro His Phe Ile Arg
                680                 685                 690
Cys Ile Lys Pro Asn Asn Ser Gln Ser Pro Gly Ile Tyr His Gln
                695                 700                 705
Gly Leu Val Leu Gln Gln Leu Arg Cys Cys Gly Val Leu Glu Val
                710                 715                 720
Val Arg Ile Ser Arg Ser Gly Phe Pro Thr Arg Met Ser His Gln
                725                 730                 735
Lys Phe Ala Arg Arg Tyr Gly Phe Leu Leu Leu Glu His Val Ala
                740                 745                 750
Ser Gln Asp Pro Leu Ser Val Ser Val Ala Ile Leu His Gln Phe
                755                 760                 765
Asp Ile Leu Pro Glu Met Tyr Gln Ile Gly Tyr Thr Lys Leu Phe
                770                 775                 780
Phe Arg Thr Gly Gln Ile Gly Lys Leu Glu Asp Thr Arg Asn Arg
                785                 790                 795
Thr Leu Asn Gly Ile Leu Arg Val Gln Ser Cys Phe Arg Gly His
                800                 805                 810
Lys Ala Arg Gln Tyr Met Lys Glu Leu Lys Arg Gly Ile Phe Asn
                815                 820                 825
Leu Gln Ala Phe Ala Arg Gly Glu Lys Thr Arg Lys Glu Phe Ala
                830                 835                 840
Ile Leu Val His Arg His Arg Ala Ala Val His Ile Gln Lys His
                845                 850                 855
Ile Lys Ala Lys Ile Ser Lys Lys Arg Phe Glu Asp Val His Gly
                860                 865                 870
Ala Thr Ile Thr Leu Gln Ala Val Ile Arg Gly Trp Leu Val Arg
```

875 880 885
Arg Cys Ser Gly Asp Ile Ala Leu Leu Gln Phe Gly Ser Gly Lys
890 895 900
Gly Asn Gly Ser Asp Glu Val Leu Val Lys Ser Ser Tyr Leu Ala
905 910 915
Glu Leu Gln Arg Arg Ile Leu Lys Ala Glu Ala Gly Leu Arg Glu
920 925 930
Lys Glu Glu Glu Asn Asp Ile Leu His Gln Arg Leu Gln Gln Tyr
935 940 945
Glu Asn Arg Trp Ser Glu Tyr Glu Leu Lys Met Lys Ser Met Glu
950 955 960
Glu Val Trp Gln Lys Gln Met Arg Ser Leu Gln Ser Ser Leu Ser
965 970 975
Ile Ala Lys Lys Ser Leu Ser Tyr Asp Asp Ser Glu Arg Asn Ser
980 985 990
Asp Ala Ser Ile Asn Thr Ala Asn Asp Glu Thr Asn Pro Pro Trp
995 1000 1005
Asp Ala Ala Thr Asn Gly Arg Arg Asn Gly Val Glu Asn Val Arg
1010 1015 1020
Pro Met Ser Ala Gly Leu Ser Val Ile Ser Arg Leu Ala Glu Glu
1025 1030 1035
Phe Glu Gln Arg Ser Gln Val Phe Gly Asp Asp Ala Lys Phe Leu
1040 1045 1050
Val Glu Val Lys Ser Gly Gln Val Glu Ala Asn Leu Asn Pro Asp
1055 1060 1065
His Glu Leu Arg Arg Leu Lys Gln Met Phe Glu Gly Trp Lys Lys
1070 1075 1080
Asp Tyr Thr Ala Arg Leu Arg Glu Thr Lys Val Ile Leu Asn Lys
1085 1090 1095
Leu Gly His Glu Asp Gly Asp Gly Glu Lys Gly Lys Lys Lys Trp
1100 1105 1110
Trp Gly Arg Leu Asn Ser Ser Arg Val Asn
1115 1120

INTERNATIONAL SEARCH REPORT

Int. tional Application No
PCT/US 99/26177

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C12Q1/68 C12N15/63 A61K38/17
C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Emest29 Database Entry Hsaa59854 Accession number AA159854; 19 January 1997 HILLIER L. ET AL.:"WashU-NCI human EST Project" XP002131112 the whole document ---	3-13
A	JAENICKE T ET AL: "THE COMPLETE SEQUENCE OF THE HUMAN BETA-MYOSIN HEAVY CHAIN GENE ANDA COMPARATIVE ANALYSIS OF ITS PRODUCT" GENOMICS,US,ACADEMIC PRESS, SAN DIEGO, vol. 8, 1 January 1990 (1990-01-01), pages 194-206, XP000653071 ISSN: 0888-7543 the whole document --- -/--	1-16,19

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 February 2000	09/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Montero Lopez, B

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte ional Application No
PCT/US 99/26177

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	LI-PING ZHAO ET AL.: "Cloning and characterization of myr6, an unconventional myosin of the dilute/myosin-V family" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 93, no. 20, 1 October 1996 (1996-10-01), pages 10826-10831, XP002131111 WASHINGTON US the whole document	1-16,19

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/26177

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claim 19 is directed to a method of treatment of the human/animal body, the search has been carried out and and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: 17, 18, 20
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17, 18, 20

Present claims 17, 18 and 20 relate to an extremely large number of possible compounds. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is not to be found, however, for any specific example of the compounds claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, no search has been carried out for claims 17, 18 and 20.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.